Marianna Sigala · Roya Rahimi ·
Mike Thelwall *Editors*

Big Data and Innovation in Tourism, Travel, and Hospitality

Managerial Approaches, Techniques, and Applications

Springer

Big Data and Innovation in Tourism, Travel, and Hospitality

Marianna Sigala · Roya Rahimi ·
Mike Thelwall
Editors

Big Data and Innovation in Tourism, Travel, and Hospitality

Managerial Approaches, Techniques, and Applications

Springer

Editors
Marianna Sigala
School of Management
University of South Australia
Adelaide, SA, Australia

Roya Rahimi
Business School
University of Wolverhampton
Wolverhampton, UK

Mike Thelwall
School of Mathemetics and Computing
University of Wolverhampton
Wolverhampton, UK

ISBN 978-981-13-6338-2 ISBN 978-981-13-6339-9 (eBook)
https://doi.org/10.1007/978-981-13-6339-9

Library of Congress Control Number: 2019930369

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

Introduction

Big Data: The Oil of the New Tourism Economy

Information has always been the lifeblood of tourism. Nowadays, technological advances related to big data further enable transformation and rapid innovation in tourism (Sigala 2018a). Technological tools enable the real time, fast, and mobile capture and sharing of a huge amount of multimedia data in a great variety of format, social media networks further facilitate the fast virality of big data fostering their enrichment, augmentation, and change. Technologies also enable the fast processing, visualization, and analyses of big data supporting and facilitating decision-making in daily operations but also for strategizing. Overall, big data has led to the creation of new technologies, methods, data capture applications, visualization techniques, and data aggregation capabilities (Gandomi and Haider 2015). In this vein, big data is traditionally described in terms of Versus: Volume, Variety, Velocity, Validity, Veracity, Value, Visibility, Visualization, Virility in spreading (Raguseo 2018; Günther et al. 2017).

Big data represent a huge opportunity, game changer, and a fuel of competitiveness and innovation in tourism. Big data can drive innovation and enhanced performance in all business operations across the business value and supply chain (Choi et al. 2017). For example, big data can enable data-driven marketing practices such as, recommendations, geo-fencing, search marketing, social Customer Relationship Marketing (CRM), market segmentation, personalization, and marketing-mix optimization (Sigala 2018b; Talón-Ballestero et al. 2018; Lehrer et al. 2018). Big data is also the major resource for developing smart tourism (Gretzel et al. 2015). Big data analytics can also enrich decision-making and market research in tourism in various areas, such as predicting tourism demand, measuring tourists' satisfaction, and designing personalised tourism experiences, destination management (Xiang and Fesenmaier 2017; Fuchs et al. 2014; Li et al. 2018; Reinhold et al. 2018). Big data does not only result in more efficient and effective operations and enhanced decision-making; big data support better strategizing and can empower tourism firms to transform their business models and strategies (Sigala 2018a).

Thus, data is considered as the “oil” of the digital economy and tourism firms need to consider and manage it as a valuable asset. However, companies have more access to big data than they know how to manage and translate it into value (Braganza et al. 2017). Little is still known about how firms can develop effective strategies for best capitalizing on big data (Wedel and Kannan 2017). Little is also known about how companies and their management should evolve to develop and implement new human skills and capabilities as well as procedures to compete in this new environment. Big data are not solely a technology issue, but rather a socio-technical issue. Hence, if firms want to make full use of big data, then they need to adopt new management mindsets, new organizational structures and cultures (e.g., cross-functional teams, corporate wide and open communication, cooperation with third parties and online platforms) as well as new work-practices such as data-driven and analytical culture.

Scope and Structure of the Book

This book brings together multidisciplinary research and practical evidence addressing the questions about the opportunities, affordances but also the challenges brought forward by big data in driving and supporting innovation in tourism. The book chapters investigate and reveal the role and application of big data in innovating and transforming tourism practices at various levels: (1) a micro-firm level and macro-destination level; and (2) strategic and operational level by showing the implementation of big data in transforming firms’ business models but also value chain operations (e.g., marketing, operations, sales, supply chain, human resource management, crisis management, smart services, smart destinations, customer experiences).

The book conceptualizes big data implementation in an input–process–outcome framework. Big data provide the inputs for transforming practices and strategies such as data sets, data sources, technological tools and devices, organizational resources, skills, and capabilities. Big data provide both the tools to support processes (e.g., big data analytics and techniques such as netnography, semantic analyses), but they also enable and foster new processes, such as: the managerial approaches (e.g., crowdsourcing, open innovation); the business operations, such as marketing, operations, supply chain, customer service, and new service development. Big data exploitation should lead to benefits to various stakeholders: customers (e.g., service, personalization); firms (e.g., performance, agility–flexibility); and societies (e.g., well-being, social value, entrepreneurship). Finally, as big data implementation happens within a broader context (PESTEL environment), big data are influenced by the context (e.g. competition, societal changes) but they also form and shape a new context (e.g. new privacy legislation, new security and intellectual property policies).

In this vein, the chapters of the book are structured around this big data process-oriented framework. The following section briefly describes the structure of the book and the contribution of the book chapters.

Content of the Book

The book starts with four chapters focusing primarily on the inputs that big data can provide. The chapters focus on two types of data inputs namely, inputs provided by Google Data Trends as well as inputs generated by the Internet of the Things (IoT) and electronic devices. The chapters discuss the features of these inputs and analyze specific examples showing the application and use of these data inputs for decision-making. The last chapter related to big data inputs develops a decision framework that users can use for evaluating and selecting inputs for big data initiatives based on various data quality criteria. Analytically, Chap. 1 is titled *Composite Indicators for Measuring the Online Search Interest by a Tourist Destination* and it is contributed by Maria Gorete Ferreira Dinis, Carlos Manuel Martins da Costa, and Osvaldo Rocha Pacheco. The authors propose a methodology for building composite indicators to measure, almost in real time, the online public interest by a tourist destination, using Google Trends data. The methodology is then applied to measure the online search interest of foreign markets, namely Spain, the UK, and Germany by Portugal as a tourist destination. Chapter 2 focusing on inputs is titled *Developing Smart Tourism Destinations with the Internet of Things* and it is written by Nicholas Wise and Hadi Heidari. This chapter discusses how the Internet of Things devices can be used to generate new tourism applications and services and how this, in turn, subsequently supports the emergence of smart cities. Josep Mª Espinet authored the chapter titled *Big Data in Online Travel Agencies and its Application Through Electronic Devices* (Chap. 3) discusses how data generated by electronic devices can help online travel agents to better understand their customers and use this insight to better manage the customer experience and services. Chapter 4 contributed by Marianna Sigala, Andrew Beer, Laura Hodgson, Allan O'Connor and titled *Big Data for Measuring the Impact of Tourism Economic Development Programmes: A Process and Quality Criteria Framework for Using Big Data* reviews the related literature and develops two frameworks that can assist big data users: a framework showing the big data processes that firms and users have to undertake for implementing big data initiatives: a decision framework identifying the data quality criteria that users can use for evaluating and selecting big data sources and sets.

The book continues with chapters focusing on the way big data advances assist tourism firms to undertake big data process. The primary focus of these chapters is on identifying and explaining various big data analytics tools and methodologies. Daniela Ferreira contributed Chap. 5 titled *Research on Big Data, VGI, and the Tourism and Hospitality Sector: Concepts, Methods, and Geographies*. The chapter conducted a bibliometric review of tourism and hospitality research publications

2011–2017 focusing on big data and Volunteered Geographic Information (VGI), which reveals the main concepts and the research methods that have been used for exploiting such data. Mike Thelwall is the author of the chapter titled *Sentiment Analysis for Tourism* (Chap. 6). This chapter discusses methods to detect the sentiment of tourists toward hotels, attractions, or resorts, as expressed in online comments or reviews about them. Extracting these sentiments gives managers a new source of automated customer feedback, allowing them to gain deeper insights into which aspects of their offerings are popular and unpopular. Chapter 7 co-authored by Konstantinos Vassakis, Emmanuel Petrakis, Ioannis Kopanakis, John Makridis and George Mastorakis lloked at methodologies for exploiting location-based data. The chapter titled *Location-Based Social Network Data for Tourism Destinations* discusses a methodology for the extraction, association, analysis and visualization of data derived from LBSNs. This provides knowledge of visitor behaviours, impressions and preferences for tourist destinations. A case study of Crete in Greece is included, based upon visitors' posts and reviews, nationality, photos, place rankings and engagement. By using data coming for two destinations (namely Heraklion and Chania, the chapters provides a case study for illustrating how the information may be visualized to reveal useful patterns for managers. Topic modeling big data strategy for analyzing text documents is the big data methodology explained by a chapter titled *Identifying Innovative Idea Proposals with Topic Models—A Case Study from SPA Tourism* (Chap. 8) and contributed by Gabriele Sottocornola, Fabio Stella, Panagiotis Symeonidis, Markus Zanker, Ines Krajger, Rita Faullant, and Erich Schwarz. The application of this methodology is explained by using a case study and data coming from spa tourism. In this case study, the documents are ideas for spa services submitted online by users and the results are compared with machine learning approaches. Steven Valcke contributed a practical case study explaining the use of big data sets and analytics for managing crisis at a destination level. The case study is entitled *Customer Data and Crisis Monitoring in Flanders and Brussels* (Chap. 9) and it shows how Visit Flanders (Belgium) has used various types of big data (including flight data, mobile data, scraping hotel review scores, and credit card data) for monitoring and managing the impacts of the terrorist attacks in November 2015 on the destination visitation and image.

The third section of the book includes chapters focusing on the outcomes of big data initiatives. George Joseph and Vinu Varghese contributed Chap. 10 titled *Analyzing Airbnb Customer Experience Feedback Using Text Mining*. The chapter shows how firms can use text mining of Airbnb user reviews to analyse and understand various aspects in order to drive customer satisfaction. Nikolaos Stylos and Jeremy Zwiegelaar authored the chapter titled *Big Data as a Game Changer: How Does it Shape Business Intelligence Within a Tourism and Hospitality Industry Context?* (Chap. 11). In their chapter, the authors show how tourism firms can use internal and external data sources for enriching their business intelligence and optimizing business processes. Irene Gil-Saura, María-Eugenia Ruiz-Molina, and David Servera-Francés co-authored a chapter titled *Strengthening Relational Ties and Building Loyalty Through Relational Innovation and Technology:*

Evidence from Spanish Hotel Guests (Chap. 12). This chapter explains how tourism firms can exploit big data for building relational capital with their customers, which, in turn, can be translated into brand equity, strong hotel–guest relational ties, and greater customer loyalty.

The book concludes with one chapter providing a wider view of the context influencing but also being shaped by big data initiatives. Mine Inanc–Demir and Metin Kozak contributed Chap. 13 titled *Big Data and its Supporting Elements: Implications for Tourism and Hospitality Marketing*. This chapter debates how big data, artificial intelligence, and IoT are likely to reshape the traditional structure of tourism and hospitality marketing in the future. The chapter also identifies and discusses the new management approaches driven by big data that are required to maintain competitiveness in a new tourism era.

Overall, of course, the book chapters do not offer a holistic view of all the big data applications, trends, and challenges. But what the book chapters offer is an in-depth discussion of the issues that they focus on, a practical application of their arguments as well as ideas and suggestions to drive future research. The variety of the book chapters also provide evidence of the multifaceted and complex nature of big data initiatives as well as of their continuously and dynamically changing and evolving aspect and environment.

We hope that you will enjoy reading this book and that you will find it inspirational to your own research but also teaching practices.

Marianna Sigala
Roya Rahimi
Mike Thelwall

References

Braganza A, Brooks L, Nepelski D, Ali M, Moro R (2017) Resource management in big data initiatives: processes and dynamic capabilities. J Bus Res 70:328–337

Choi TM, Chan HK, Yue X (2017) Recent development in big data analytics for business operations and risk management. IEEE Transac Cybernet 47(1):81–92

Fuchs M, Höpken W, Lexhagen M (2014) Big data analytics for knowledge generation in tourism destinations–A case from Sweden. J Destinat Mark Manage 3(4):198–209

Gandomi A, Haider M (2015) Beyond the hype: big data concepts, methods, and analytics. Int J Informat Manage 35(2):137–144

Gretzel U, Sigala M, Xiang Z, Koo C (2015) Smart tourism: foundations and developments. Electron Mark 25(3):179–188

Günther WA, Mehrizi MHR, Huysman M, Feldberg F (2017) Debating big data: a literature review on realizing value from big data. J Strategic Informat Syst 26:191–209

Lehrer C, Wieneke A, vom Brocke J, Jung R, Seidel S (2018) How big data analytics enables service innovation: materiality, affordance, and the individualization of service. J Manage Informat Syst 35(2):424–460

Li J. Xu L, Tang L, Wang S, Li L (2018) Big data in tourism research: a literature review. Tourism Manage 68:301–323

Raguseo, E (2018) Big data technologies: an empirical investigation on their adoption, benefits and risks for companies. Int J Inf Manage 38(1):187–195

Reinhold S, Laesser C, Beritelli P (2018) The 2016 St. Gallen consensus on advances in destination management. J Destinat Market Manage 8:426–431

Sigala M (2018a) New technologies in tourism: from multi-disciplinary to anti-disciplinary advances and trajectories. Tourism Manage Perspect 25:151–155

Sigala M (2018b) Implementing social customer relationship management: a process framework and implications in tourism and hospitality. Int J Contemp Hospital Manage 30(7):2698–2726

Talón-Ballestero P, González-Serrano L, Soguero-Ruiz C, Muñoz-Romero S, Rojo-Álvarez JL (2018) Using big data from customer relationship management information systems to determine the client profile in the hotel sector. Tourism Manag 68:187–197

Wedel M, Kannan PK (2016) Marketing analytics for data-rich environments. J Mark 80 (6):97–121

Xiang Z, Fesenmaier DR (2017) Big data analytics, tourism design and smart tourism. In: Analytics in smart tourism design, Springer, Cham, pp 299–307

Chapter 1
Composite Indicators for Measuring the Online Search Interest by a Tourist Destination

Maria Gorete Ferreira Dinis, Carlos Manuel Martins da Costa and Osvaldo Manuel da Rocha Pacheco

Abstract This chapter presents a methodology for building composite indicators to measure the public online search interest by tourist destinations. As an example, we have applied it to measure the online search interest of foreign markets, namely Spain, the UK and Germany by Portugal as a tourist destination. In order to build the composite indicators we extracted weekly and during one year, data from the Google Trends (GT) tool, based on the set of search terms chosen to define the destination Portugal. The composite indicators proposed are based on the Tourism Satellite Accounts (TSA) conceptual framework and weighted by the arithmetic mean of seven primary indicators composed by fifteen sub-indicators. The results indicate the interest and popularity of Spanish, British and German foreigners by tourism in Portugal and country specific touristic products. The obtained results contribute definitively to support and help Destination Management Organizations (DMO) enabling timely decisions.

Keywords Composite indicators · Search interest · Google Trends · Portugal

1.1 Introduction

For a tourist destination to be sustainable and competitive it must be managed by an organization, that regardless of its nature, should perform several functions. We emphasise in this study the role in the marketing of the destination and the duty to produce and disseminate information (Ritchie and Crouch 2003). On the other hand, public policies, particularly at European Union level, point to the need for a sustainable development strategy for tourism in Europe, which among other

M. G. F. Dinis (✉)
GOVCOPP, Polytechnic Institute of Portalegre, Portalegre, Portugal
e-mail: gdinis@ipportalegre.pt

C. M. M. da Costa
GOVCOPP, University of Aveiro, Aveiro, Portugal

O. M. da R. Pacheco
IEETA, University of Aveiro, Aveiro, Portugal

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_1

measures, mentions the need for methodologies and indicators for monitoring the tourism sector (CCE 2001).

The National Tourism Authority in Portugal recognizes in the national strategic plan for tourism, (2013–2015) the need to align the communication strategy of the destination with the new trends of the sector, maintaining a relation of proximity with the potential consumer, betting on the Internet as a channel of communication and of the presence in the different stages in the decision process. To this end, it reinforces the need to use tools to analyse consumer behaviour and trends on the Internet and to monitor the results obtained by investing in digital marketing metrics (Resolução do Conselho de Ministros n.º 24/2013 de 16 de abril).

Consumer behaviour in tourism has changed significantly in recent years, the Internet being one of the major drivers of these changes. According to Jeng and Fesenmaier (2002) in Leung et al. (2013), the traveller uses the Internet at the beginning of the decision-making process in order to make the right decisions. Currently, the consumer uses the Internet in all phases of the travel cycle, from the dream phase to the sharing of experiences.

Usually the journey's decision-making process begins with the search of information in the search engines, Google being the most used worldwide (StatCounter Global Stats 2018a). Due to this fact, the GT tool shows almost in real time, the individual's interests, according to certain topics, based on the searches performed. This information presents a great potential for the knowledge and understanding of the future consumer needs in tourism.

The DMOs need useful and timely information to assist them in the decision making process, which is increasingly done timely and with shorter planning times. Aligned with this need, the present chapter aims to propose a methodology for building composite indicators to measure the online public interest, and to show that the data made available through the GT tool can be used to know almost in real time and much earlier than the official statistical data, the interest of the tourist by Portugal. Our results provide a groundwork for DMOs to analyse the daily online public interest by tourism destination and products, helping them to make informed and timely decisions.

1.2 Literature Review

1.2.1 Tourism and Statistical Information

Ritchie and Crouch (2003) state that for a destination to be competitive and sustainable it must be managed by an organization that should perform the development of effective marketing channels that facilitate the connection between the destination and the potential consumer, as well as strategically selecting the potential target markets for the destination. In addition, the organization must collect and manage the information for internal use and distribute it to tourism stakeholders. Among the information advised, we highlight the information about visitor behaviour patterns and the monitoring of target markets. Buhalis and Amaranggana (2015) argue that the competitiveness of a destination may increase by applying a smartness concept to understand and address the travellers' needs, wishes and desires before, during and after their travel.

The smartness concept relates, in the opinion of Höjer and Wangel (2015), not so much with the technological advances, products or services per se but with the interconnection, synchronization and work concerted of these. On the other hand, Gretzel et al. (2015) refer to the concept of smart tourism as the capacity of intelligently storing, processing, combining, analysing and using big data to inform business innovation, operations and services. By applying this concept to tourism destinations, Buhalis and Amaranggana (2015) say that smart tourism destinations should make optimal use of the large data sets known as Big Data, coming from the instant information exchange, to know patterns and trends and to offer the right services that suit the users' preferences at the right time, contributing this way to enhance their tourism experience.

Tourism is a sector with unique characteristics, highlighting the intangibility and perishability of the products and the susceptibility to natural, economic and political phenomena, therefore it requires large volumes of intensive, updated, timely and relevant information, to support and help the decision-making process.

In Portugal the statistical information is collected, analysed and made available to tourism organizations by the National Institute of Statistics (INE). In relation to inbound tourism, the indicators obtained result from the application of monthly surveys of tourist lodgings with the purpose of knowing the offer and occupation of these same lodgings, with no information on foreign tourists staying in other accommodation facilities (for example, friends and family) or hikers visiting Portugal. Moreover, INE displays the indicators only monthly and the first results are made available to the public forty three days after the reference period. The final results are published annually around seven months later, which allows us to conclude that, in addition to being limited, statistical information is made available to tourism organizations very late (Dinis 2016b).

1.2.2 Consumer Behaviour in Tourism, the Internet and Google Trends

Consumer travel's consumption patterns and behaviour in tourism have changed significantly. Nowadays, the consumers of tourism are more informed, experienced, technologically able, more independent and more involved (Poon 1993; Buhalis et al. 2006), and gradually the consumer is being placed in the sector's driving seat (Buhalis et al. 2006), becoming the central player in the process of creating and shaping brands and experiences (Gretzel et al. 2006). Travellers increasingly use ICT throughout all phases of their travel, this means that consumers are also more engaged

(Gretzel et al. 2006), using ICT not only to obtain information but also to create and share content and explicitly express their opinions and points of view (Gurău 2008).

The users generated comments (UGC) in review sites or in social networks are perceived, from the point of view of the potential consumer, as an important source of information, more reliable than official sources due to their width and depth (Milano et al. 2011), but also one of the key marketing tool for DMOs, since UGC increasingly influenced destination awareness and decision-making to select the tourist destination (Tussyadiah and Fesenmaier 2009).

Gretzel et al. (2006) reported that the new role and proactive attitude of the consumer requires the development of appropriate marketing strategies consumer-centric, which have to be defined, not in terms of rigid socio-demographic characteristics, but in terms of the dynamic preferences. Gurău (2008) added that given the interactive dimension of the Internet, the organizations shall analyse the direct and the indirect feed-back transmitted by relevant audiences connected to the Internet. Moreno de la Santa (UNWTO 2011) founded that marketers have focused their strategic actions more on the reservation phase and underestimated the importance of the dream, search and experience phases, essential to influencing the decision making and consumer loyalty.

Cox et al. (2009) reported that research conducted to date demonstrated that consumers use different types of online information sources depending on the travel planning process phase. Xiang and Fesenmaier (2006) supplemented that even if there exists multiple online information sources where travellers can find the information they need, it is evident that a large proportion of tourism consumers start the information search process via an Internet search engine. Gretzel et al. (2006) posited that face to the enormity and variety of information existing on the Internet, the statistical evidence shows that the consumers depend strongly on the search engines to find the desired information. In the study carried out by Dinis et al. (2016a), the authors verified that the use of online information sources during the travel decision-making process varies, not only at the stage of the travel planning process but also depending on the country's origin of the consumer, and it has been found that the results of search engines have influence throughout all the decision-making process, although there is a primacy in the pre-purchasing phase, which is most evident for consumers in Japan, the USA, Germany and France.

According to data from StatCounter Global Stats (2018b), the Google search engine in Europe, between January 2009 and December 2017, had a market share above 91%. This fact confirms that the data regarding searches performed in Google may be considered representative when analysing the tendencies and intentions of the consumers on a certain subject that must be explored and used as the basis for the decision-making.

Currently, there are several tools available in the market to perform web analytics. Organizations can use one or more tools at the same time and their choice depends, among other factors, on the organization's needs and what it intends to measure (UNWTO and ETC 2008; Jackson 2009; Kaushik 2010).

The GT is a free of charge tool available in the market and launched by Google in 2012, that has incorporated Google Insights for Search (GIS), which provides search volume statistics for selected search terms, over specific time ranges and geographic regions, on a daily or weekly basis, based on patterns of queries conducted on the Google search engine. The GT classifies the searches performed in Google, according to the subject, in certain categories. The searches related to travel have been classified in a specific category with that designation, which subsequently is subdivided into eleven subcategories. Since its launch, GT data has been used by researchers in several areas of knowledge, namely health, economy, finance, communication and marketing and tourism. In the field of tourism's studies, several researchers (Chamberlin 2010; Choi and Varian 2009; Shimshoni et al. 2009; Suhoy 2009; Smith and White 2011; Artola and Galán 2012; Saidi et al. 2010; De La Oz Pineda 2014; Dinis et al. 2015; Gawlik et al. 2011; Shimshoni et al. 2009; Jackman and Naitram 2015; Pan et al. 2012; Li et al. 2017) have shown in their work the existence of similarities between the GT data and the tourism official statistical data and the potential of the GT data to help in the tourism demand prediction of a given destination. However, to the best of our knowledge, this is the first study that uses the GT data for constructing composite indicators that measure the online interest of a certain country's tourism.

1.2.3 Indicators and Web Analytics Strategy

The conceptualization of indicators has been addressed by several authors and public organizations (Monjardino 2009; Gahin et al. 2003 in White et al. 2006; Bossel 1999; OECD 2003; EEA 2005; SREA et al. 2006; EC 2006). It can be said that the authors are unanimous in considering the indicators as a measure that results from direct observation or analysis of basic information and that provides information that assists understanding a given phenomenon. According to the number of variables involved, the indicators can be distinguished between simple or analytical indicators, when they are constituted by only one variable; and composites, synthetic or indices, when they result from a composition of variables (Castro Bonaño 2002).

Composite indicators are indicators used to measure complex phenomena in a simple way, facilitating communication with the public and are suitable to compare different territorial domains (Segnestam 2002 in Hugony and Cladera 2008), a fact that justifies the adoption of this type of indicator in the empirical part of this study.

On the other hand, web analytics refers to "the measurement, collection, analysis and presentation of Internet data in order to understand and optimize the use of the Web" (UNWTO and ETC 2008, p. 1). However, Kaushik (2010) found that organizations have limited the use of web analytics to only analysing visitor data on a particular website. In the opinion of the same author, one way organizations have to gain strategic advantage is to include competitive intelligence data in its web analysis strategy, ensuring that the organization's decision-making is grounded with information not only related to the organization's performance, but also considering the competitors performance or the industry in general, recommending the GT and GIS tools to obtain this type of data, namely for analysing online search behaviour and audience targeting.

1.3 Methodology

The aim of this study is to show how to build composite indicators to measure the public online search interest by tourist destinations. We applied it to measure the online search interest of foreign markets, namely Spain, the UK and Germany by Portugal as a tourist destination.

The proposed composite indicators are generically called *Google Output Relevance Indicator External* [GORE (country of residence)_PT: TOURISM]. They are simple and easy to understand and are available daily, allowing organizations, especially DMOs, to obtain up-to-date information about the interest and popularity of a certain tourism destination over a specific population.

There is no methodological procedure for the construction of synthetic indicators that is unique or the most appropriate, and it must be selected based on the specific needs (Pérez et al. 2009). Based on the work of other authors (UNWTO 2004; OECD 2008), the derivation of composite indicators shall have the following phases: (i) theoretical framework; (ii) selection of primary indicators; (iii) selection of search terms and geographical locations in GT; (iv) transformation, weighting and aggregation of primary indicators; and (v) validation and reliability of indicators.

The reliability of the indicators was analysed using the Statistical Package for the Social Sciences (SPSS version 20).

1.3.1 Theoretical Framework

The existence of a theoretical framework is fundamental in the construction of composite indicators. The framework provides the basis for the selection and combination of variables into a meaningful composite indicator and should clearly define the phenomenon to be measured. The theoretical framework can be defined appealing to experts or stakeholders in the area. Ultimately, the composite indicators users should assess its quality and relevance (OECD 2008).

In this study, the theoretical framework adopted as the structuring axis for the construction of the composite indicators was the TSA model, an international standard methodological framework for organizing statistical data on tourism adopted in several countries, including Portugal. In this way, we considered the tourism characteristic products identified in the TSA model. This is to say, the products which probably cease or where consumption would reduce significantly without tourism, as being the products that are the nucleus of the tourist activity and that can be the origin of the interest of the potential consumer by a tourist destination.

1.3.2 Selection of Primary Indicators

The next step in the creation of the indicators is the selection of the primary indicators that integrate the composite indicators. This process should be based on the primary indicators evaluation in relation to basic criteria, defined from the characteristics of the indicators, and listed according to their apparent utility (UNWTO 1996). In this study, we have considered as relevant the following characteristics of the primary indicators: (i) relevance and representative of the indicators for the explanation of the tourism phenomenon; (ii) comparability of the indicators over time and across countries; (iii) data available; and (iv) simple and easy to understand.

As can be observed in Table 1.1, we have selected 15 primary indicators, in terms of search volume index (SVI) of the GT, as being representative of the online interest of potential consumers for the characteristic products of tourism in Portugal. Due to the diversity of the products that integrate “cultural services” and “sports and recreational services”, we have chosen to assign four and three primary indicators, respectively, that in our opinion best represent the product, taking into account the existing classification and categories in the GT tool.

1.3.3 Selection of Search Terms and Geographical Locations in GT

The composite indicators intend to show daily the online interest of tourism in Portugal. Considering that, some authors (Pan et al. 2006; Sanderson and Kohler 2004; Jones et al. 2008), conclude that the searches related to travelling information in search engines integrates as keywords the designation of the city and/or country, being that, many times these keywords are accompanied by others referring to the aspects of the trips (i.e. attractions, transports and restaurants).

In order to represent Portugal it was considered as search terms the name of the country, the name of the regional areas of tourism,[1] except the “north” and the “center” because based on our own search experience they are ambiguous search terms, and the name of the municipalities,[2] with the greatest number of overnight stays from the country of residence under analysis, according to official statistics (INE 2012a, b, c, d, e, f, g). In addition and where appropriate, it was considered as search term the designation of relevant tourist resources in the municipality, since the municipality has already been considered by the previous criterion. The search terms are grouped in a single entry in GT using the plus sign. The quotation marks are used when we want to considerer searches that match exactly that municipality (e.g. “porto”).The minus sign was used when we wanted to exclude search terms that

[1]In mainland Portugal the regional tourism areas are: Porto and the North, Center of Portugal, Lisbon, Alentejo/Ribatejo and Algarve.

[2]The municipality was used because there are no overnights data with disaggregation at city/local level.

Table 1.1 Primary indicators framework

Tourism characteristic produts	GT category	GT subcategories/primary indicators	Abbreviations
Food and Beverage Services	Food and Beverage	Restaurants	RESTAUR
Accommodation Services	Travel	Hotels and accommodation	HTALOJ
Passenger Transport Services and Transport Equipment Rental Services	Travel	Air travel	VAEREA
	Travel	Buses and trains	AUTCOMB
	Travel	Cruises and charters	CRUZECH
	Travel	Rental car and taxi services	RENTACAR
Travel Agencies and other Reservation Services	Travel	Travel agencies: holidays offer	AVFERIAS
Cultural Services	Reference	Zoos-aquariums-reservations	JARDZOO
	Travel	Historical sites and buildings	EDIFHIST
	Travel	Library and museums	BMUSEU
	Arts and Entertainment	Concerts and music festivals	CFESTIV
Sports and Recreational Services	Travel	Thematic parks	PTEMATIC
	Travel	Mountain resorts and ski	MONTSKI
	Sports	Golf	GOLFE
Miscellaneous Tourism Services	Travel	Beaches and islands	PRAIA

can negatively influence the results (e.g. mare, because there is a hotel in Portugal with the designation Hotel Porto Mare). The criterion to exclude these search terms is a hint for the top searches by GT.

In relation to the geographical locations, it is proposed to construct composite indicators for the main tourism markets for Portugal, namely Spain (GORE (ES)_PT: TOURISM), United Kingdom (GORE (UK)_TOURISM) and Germany (GORE (DE)_PT: TOURISM). We choose these countries because they represent approximately 50% of the total number of foreign overnight stays in Portugal (INE 2012c). In Fig. 1.1, we can observe the search terms selected for the composite indicators.

1.3.4 Transformation, Weighting and Aggregation of Primary Indicators

Before aggregating the primary indicators, the data should be analysed in order to identify and treat atypical values and/or missing cases, and to verify whether it is necessary to process the data and thus, ensuring that the conditions are met to apply the intended statistical techniques.

In this study, no transformation of primary indicators was carried out because the indicators are all in the same unit of measure and have already been normalized and scaled by Google. In addition, we kept the extreme values and outliers because the values in this context have a meaning, since the 0 in GT is showed when the search volume is low and the maximum value assumed is 100 and means the peak popularity of the search interest. We decide that primary indicator assume also the figure "zero" when there was not enough research volume to generate data, so no primary indicator is eliminated, allowing comparisons between the composite indicators proposed in the study.

The method used for weighting the primary indicators was the arithmetic mean, mainly due to the its simplicity, which means that all primary indicators have the same weight. In the following equation, we can see that each composite indicator results from the aggregation of 7 primary indicators that correspond to the 7 tourism characteristic products identified in the TSA model (Table 1.1), each with a weighting of 1/7. In the case of the primary indicators: "Passenger Transport Services and Transport Equipment Rental Services", "Cultural Services" and "Sports and Recreational Services", to avoid double counting, the value of the primary indicator results from the arithmetic mean of the indicators that constitute it. Formally,

$$
\begin{aligned}
&1/7(RESTAUR) + 1/7(HTALOJ) + 1/7[1/4(VAEREA) + 1/4(AUTCOMB) + \\
&\quad 1/4(CRUZECH) + 1/4(RENTACAR)] + 1/7(AVFERIAS) + 1/7[1/4(JARDZOO) + \\
&\quad 1/4(EDIFHIST) + 1/4(BMUSEU) + 1/4(CFESTIV)] + 1/7[1/3(PTEMATIC) + \\
&\quad 1/3(MONTSKI) + 1/3(GOLFE)] + 1/7(PRAIA).
\end{aligned}
$$

SPAIN	portugal+lisboa+alentejo+algarve+oporto+albufeira+cascais+ourem+fatima+portimao+coimbra+aveiro+tavira+gaia+douro+loule+setubal+sintra+braga+almada+evora+matosinhos+faro+varzim+guimaraes+"viana do castelo"+sesimbra+bispo
UNITED KINGDOM	portugal+lisbon+alentejo+algarve+albufeira+loule+portimao+"porto"-seguro+oporto+carvoeiro+douro+cascais+tavira+faro+montegordo+sintra+silves+evora+coimbra+almada+varzim+gaia+oeiras+matosinhos+bispo+braga+sesimbra+ourem+guimaraes
GERMANY	portugal+lissabon+lisboa+alentejo+algarve+albufeira+portimao+"porto" - seguro-mare+douro+cascais+montegordo+loule+sintra+tavira+evora+silves+faro-jandia+gaia+ourem+fatima+sesimbra+bispo+matosinhos+setubal+braga+aveiro+mafra

Fig. 1.1 Search terms of the composite indicators

1.3.5 Data Collection from GT

The composite indicators proposed have a daily character, but the primary indicators data was obtained weekly by the authors from GT, and refers to the searches carried out in Google in a wider period "the last 90 days". We opted for this time range because for a shorter time period (past 7 or 30 days), the SVI are lower, with less data available in GT.

The authors collected the data, every Saturday, for one year (March 24, 2013 to March 23, 2014). When the SVI for certain primary indicators was not enough, GT presented the data per week, so in such situations we considered that the SVI per day was equal to the SVI indicated for that week. Moreover, we choose to extract the data from Google web search in the categories "Travel", Food and Beverage", "Reference", "Arts and Entertainment", and "Sports" do GT (Table 1.1). The GT data was used in this study because Google is the most used search engine in the world, and also because the GT is, together with Baidu index, the most popular web search data used in tourism research (Li et al. 2018).

1.3.6 Validation and Reliability of Indicators

The quality of the indicators can be evaluated through two of its main characteristics: validity, and reliability. Indicators are valid if they are scientifically generated; provide relevant information; are useful and used by decision-makers (Bockstaller and Girardin 2003); and reliable if, the measures used to measure the phenomenon are consistent, that is, independent of the analyst who measures it, and whose results are repeated in consecutive measurements.

The validation is a necessary procedure for evaluate the quality of the indicators. However, few authors address this issue and propose a detailed methodology for the validation of the indicators (Bockstaller and Girardin 2003). In this study we considered the methodology of Carmines and Zeller (1979) that refers to the existence of three basic types of validation: criterion-related validity; content validity; and construct validity. Regarding the criterion-related validity, the proposed indicators were validated by relating them to other similar known indicators, in this case, the indicators that most closely resembles the proposed indicators is the SVI obtained in the GT for the category "travel", whose sample was collected, weekly, in the same time period and following the same methodological criteria as the primary indicators.

Since the composite indicators are based on the TSA conceptual framework, we have considered not necessary to perform a content validity and construct validity, as suggested by Carmines and Zeller (1979).

1.4 Results

1.4.1 Search Interest of Foreign Markets for Tourism in Portugal

In this section we present the results obtained from the calculation of the proposed composite indicators to measure the search interest of foreign markets, namely Spain (GORE (ES)_PT: TOURISM), United Kingdom (GORE (UK)_PT: TOURISM) and Germany (GORE (DE)_PT: TOURISM) for tourism in Portugal.

Analysing Fig. 1.2 it can be observed that the search interest of the Spanish and British by tourism in Portugal presents between March and August some fluctuations, however, the popularity of the country assumes values above 40. Between the months from September to December 2013 there is a decrease in interest, which is more pronounced among the Spanish. From January to March 2014 the popularity of tourism in Portugal, in general, reaches slightly lower values than the period from May to July. The GORE (DE)_PT: TOURISM indicator shows similar behaviour, however, it should be noted that this indicator almost always shows the lowest peaks of interest between May until September and between January until March 2014. In addition, the interest of German by tourism in Portugal never exceeds the value 70.

The GORE (UK)_PT: TOURISM and GORE (ES)_PT: TOURISM indicators showed the maximum interest of 82.1 on the 23rd and 24th of June 2013, respectively. On the other hand, GORE (DE)_PT: TOURISM reached the maximum value of 70.3 on the 7th of April 2013. It is important to note that the GORE (DE)_PT: TOURISM and GORE (UK)_PT: TOURISM also shows high peaks of interest, close to maximum values, in March, April and January, corroborating the conclusions pointed out by Rheem (2012) that consumers in the United Kingdom and Germany are the ones who start in advance to plan their travel on the Internet.

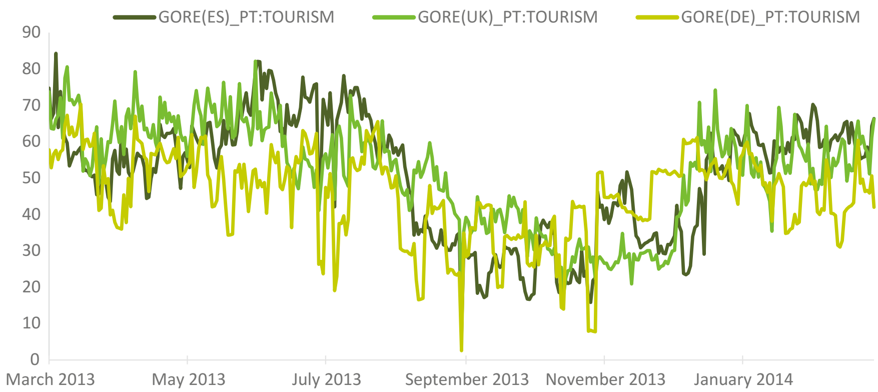


Fig. 1.2 Search interest indicators for tourism in Portugal

Regarding the minimum value, the lowest values are reached in 22nd of September by the GORE (DE)_PT: TOURISM (2.5), followed by GORE (ES)_PT: TOURISM (6.7), and lastly by GORE (UK) TOURISM (17.4).

Analysing Table 1.2, it can be seen that the GORE (UK) _PT: TOURISM average is 51.4, higher than the GORE (ES) _PT: TOURISM (50.4) indicator and the GORE (DE) _PT: TOURISM (45.1), however, GORE (ES) _PT: TOURISM is the indicator that shows the highest standard deviation (16.6). The most frequent (mode) in the GORE (UK) _PT: TOURISM is 52.3, while in GORE (ES) _PT: TOURISM is 34.3, and in GORE (DE) _PT: TOURISM is 40.7. This shows that the British have a more regular interest in tourism in Portugal over the period under analysis and, on average, higher than in other countries.

Analysing the primary indicators that gave rise to each indicator, it was found that the SVI on libraries and museums (BMUSEU) and mountain and ski resorts (MONTSKY) in GORE (UK)_PT: TOURISM, Golf (GOLFE) in GORE (ES)_PT: TOURISM, and concerts and festivals (CFESTIV) and theme parks (PTEMATIC) in GORE (DE)_PT: TOURISM assume the null value because the SVI is lower for the GT generate data (Table 1.3).

In addition, we observed that the most popular tourism products, on average, among the British are: restaurants (RESTAUR); historic buildings (EDIFHIST); cruises and charters (CRUZECH); buses and trains (AUTCOMB); hotels and accommodation (HTALOJ); and air travel (VAEREA). On the other hand, Germans show a greater interest in rental car (RENTACAR), travel agencies/holiday offer (AVFERIAS), zoos, aquariums and reservations (JARDZOO), buses and trains (AUTCOMB), restaurants (RESTAUR) and historical buildings (EDIFHIST); and the Spanish present, on average, higher SVI on mountain resorts and sky (MONTSKY), holiday offer (AVFERIAS), buses and trains (AUTCOMB), air travel (VAEREA), concerts and festivals (CFESTIV), hotels and accommodation (HTALOJ) and beaches (PRAIA).

Comparing the composite indicators, it can be seen that there are tourism products that are more popular in certain markets than in others, such as mountain and ski resorts (MONTSKI) and concerts and festivals (CEFESTIV) which are more popular among the Spanish; rental car and taxi services (RENTACAR) and holiday offerings

Table 1.2 Descriptive statistics of external search interest indicators by tourism in Portugal

Statistics	GORE (UK)_PT:Tourism	GORE (ES)_PT:Tourism	GORE (DE)_PT:Tourism
Average	51.4	50.4	45.1
Standard deviation	14.3	16.6	12.0
Mode	52.3	34.3	40.7
Minimum	17.4	6.7	2.5
Maximum	82.1	82.1	70.3
Observations	365	365	365

Table 1.3 Average of primary indicators by indicator of external search interest in tourism in Portugal

Indicators	Primary indicator average							
	RESTAUR	CFESTIV	GOLFE	BMUSEU	AVFERIAS	RENTACAR	AUTCOMB	CRUZECH
GORE (UK)_PT:Tourism	73.1	30.2	54.0	0.0	44.8	51.2	67.7	66.1
GORE (DE)_PT:Tourism	61.7	0.0	10.7	0.0	63.3	68.8	62.7	50.6
GORE (ES)_PT:Tourism	44.1	57.4	0.0	36.3	63.5	18.0	63.0	60.1

Indicators	Primary indicator average						
	MONTSKI	JARDZOO	EDIFHIST	PTEMATIC	PRAIA	HTALOJ	VAEREA
GORE (UK)_PT:Tourism	0.0	61.6	70.3	24.2	55.4	59.1	58.2
GORE (DE)_PT:Tourism	2.3	63.3	59.0	0.0	48.8	47.0	56.4
GORE (ES)_PT:Tourism	65.0	55.9	43.2	41.4	55.4	55.8	61.1

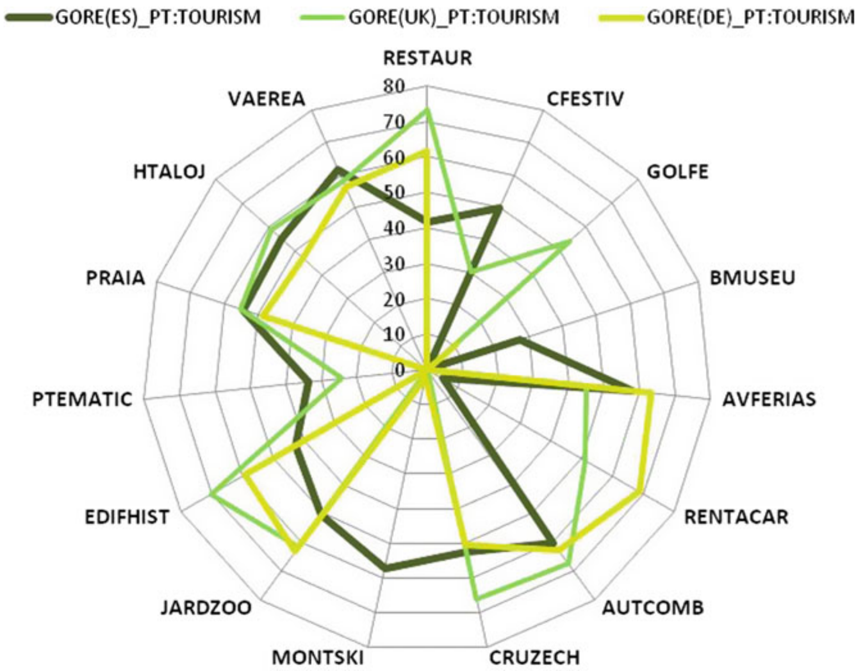


Fig. 1.3 Graphical representation of the mean of the primary indicators, by composite indicator

(AVFERIAS) are the subject of greater search interest by the Germans; cruises and charters (CRUZECH), historic buildings (EDIFHIST), golf (GOLFE) and restaurants (RESTAUR) are the most searched, on average, by the British (Fig. 1.3).

1.4.2 Validation and Reliability of Composite Indicators

The composite indicators proposed in this study were validated in relation to the criterion-related validity. Thus, the composite indicators were correlated with the SVI obtained in the GT for the category "travel".

By analysing Table 1.4, we verified that the proposed composite indicators, namely GORE (UK)_PT: TOURISM and GORE (ES)_PT: TOURISM have a high correlation with the SVI on "travel", with Pearson coefficients close to the unit, which means that there is a major competition between the proposed indicators and the criterion indicator. The GORE (DE)_PT: TOURISM is the indicator that shows the lowest correlation coefficient (0.3).

Regarding the indicators reliability or consistency, these were analysed through the Cronbach's Alpha. Analysing Table 1.5, we conclude that the GORE (ES)_PT: TOURISM and GORE (UK)_PT: TOURISM indicators are the most reliable because

Table 1.4 Pearson correlation coefficient between composite indicators and SVI on "travel"

Composite indicators	SVI "travel"
GORE(ES)_PT:TOURISM	0.8[a]
GORE(UK)_PT:TOURISM	0.9[a]
GORE(DE)_PT:TOURISM	0.3[a]

[a]The correlation is significant at the 0.01 level (2 extremities)

Table 1.5 Reliability of the indicators, according to Cronbach's alpha value

Indicators	Alfa de Cronbach value	Cronbach Alpha based on standardized items	Number of items
GORE(ES)_PT:TOURISM	0.79	0.80	14[a]
GORE(UK)_PT:TOURISM	0.81	0.84	13[a]
GORE(DE)_PT:TOURISM	0.66	0.69	12[a]

[a]SPSS has removed the indicator(s) from the analysis that have a "zero" variance

they are those with a Cronbach Alpha nearest to the unit. In addition, alpha values indicate that the items (indicators) of the scale are inter-correlated, considering that, alpha values above 0.7 are satisfactory and above 0.8 are good (Hill and Hill 2002). Therefore, the primary indicators, excluding those indicators that presented "zero" variance and were not included in the analysis, are generally important for the computation of the respective composite indicators.

1.5 Discussion

The results of this exploratory study show that the interest by tourism in Portugal presents a seasonal behavior and quite satisfactory levels of interest in much of the time period under analysis, which indicates a possible increase of the inbound tourism in Portugal, that in general has occurred in the recent years. The composite indicators show similarities with the effective tourist demand of these markets for tourism in Portugal, similar to the study realized by the authors (Dinis et al. 2016a), proving to be a proxy indicator of inbound tourism in Portugal. However, it is not possible to carry out this comparative study because the GT data used in the construction of the indicators, refers to the searches carried out by the users in the last 90 days and, in addition, the official indicators on inbound tourism in Portugal are presented by month, a gap identified in the literature review and that contributed also for the realization of this work. In this study, due to the limitations of GT, we have selected a restricted number of search terms to represent Portugal, however, we recognize that the results could be different if we choose other search terms or analyse the search interest by tourism region, as evidenced by the research carried out by the author (Dinis et al. 2016b). In future research, it would be interesting to construct composite indicators to measure the search interest by these markets for different touristic destinations in Portugal. In addition, the construction of composite indicators considering more recent data, would allow for the comparison of results, to verify if there were significant changes in online popularity of Portugal as tourism destination along time.

1.6 Conclusion

The objective of this study was to present a novel methodology for building composite indicators to measure daily the online public interest by tourist destinations. In this exploratory study, we applied the methodology to measure the interest of Portugal's main tourism foreign markets, namely Spain, United Kingdom and Germany. The main reason for this study was the need to develop indicators made available timely to tourism organizations that reflect the behaviour and intentions of the potential consumer in relation to the tourism sector. The consumer currently uses the Internet throughout the travel decision-making process, not only to find information or to make a reservation, but also to share the experiences, express opinions and interact with the public. The consumers activity on Internet search engines leaves a digital footprint, that in case of Google are made available via the GT tool. To construct the indicators, the authors collected the data from GT, weekly, during one year, following a methodology, to the best of our knowledge, never proposed by others authors.

The results reveal that foreign search interest by the tourism in Portugal decreases between the months of September to December of 2013, mainly between the Germans and the Spanish. On average, Internet users in the UK show a greater interest of tourism in Portugal than individuals from other foreign countries. The results concerning the interest by product characteristics of tourism shows that this varies according to the origin of the individuals, with restaurants and historic buildings being the most popular products among United Kingdom internet users, ski resorts are of greater interest by the individuals of Spain and the rental car by the Germans.The proposed indicators were validated and their reliability tested, presenting Cronbach's alpha values considered good for the GORE (ES)_PT: TOURISM and GORE (UK)_PT: TOURISM and satisfactory for GORE (DE)_PT: TOURISM.

This study helps to understand the behaviour and interests of the potential consumer segmented by a geographic criteria regarding tourism in Portugal, being of great importance for tourism organizations, namely for DMOs. It allows obtaining information on a regular basis, which helps in decision-making, particularly in delineating and developing the online marketing strategy.

References

Artola C, Galán E (2012) Tracking the future on the web: construction of leading indicators using Internet searches. Banco de Espana Occasional Paper, (1203). http://bit.ly/XRLfCf. Accessed 25 Nov 2012

Bockstaller C, Girardin P (2003) How to validate environmental indicators. Agric Syst 76(2):639–653

Bossel H (ed) (1999) Indicators for sustainable development: theory, method, applications. International Institute for Sustainable Development, Winnipeg

Buhalis D, Amaranggana A (2015) Smart tourism destinations enhancing tourism experience through personalisation of services. In: Tussyadiah I, Inversini A (eds) ENTER 2015 Proceedings, Lugano. Springer, Wien

Buhalis D, Costa C, Ford F (2006) Tourism business frontiers: consumers, products and industry. Butterworth-Heinemann/Elsevier, Burlington

Carmines EG, Zeller RA (1979) Realiability and validity assessment. Quantitative applications in the social sciences. Sage Publications, EUA

Castro Bonaño DJM (2002) Indicadores de desarrollo sostenible urbano: una aplicación para Andalucía. Tesis doctoral, Universidad de Málaga, Málaga

Chamberlin G (2010) Googling the present. Econ Labour Market Rev 4(12):56. http://bit.ly/UXfbLB. Accessed 16 Nov 2012

Choi HC, Varian H (2009) Predicting the present with Google Trends. http://bit.ly/1zODvPz. Accessed 12 Nov 2011

Comissão das Comunidades Europeias [CCE] (2001) Uma abordagem cooperativa para o futuro do turismo europeu. (COM (2001) 665 final) de 13.11.2001), Bruxelas

Cox C, Burgess S, Sellitto C, Buultjens J (2009) The role of user-generated content in tourists' travel planning behavior. J Hosp Market Manag 18(8):743–764

De La Oz Pineda ME (2014) Preditictingtourist inflows to Punta Cana, Dominican Republic, Using Google Trends. All Graduate Plan B and other Reports, 360. http://bit.ly/1nWwoEk. Accessed 20 Dec 2015

Dinis G, Costa C, Pacheco O (2015) Nós Googlamos! Utilização da ferramenta Google Trends para compreender o interesse do público pelo Turismo no Algarve. Dos Algarves: Multidisc e-J 26(1):64–84

Dinis G, Costa C, Pacheco O (2016a) The use of Google Trends data as proxy of foreign tourist inflows to Portugal. Int J Cult Digital Tour 3(1) (Spring)

Dinis G (2016b) Indicadores de comportamento online e tendências da procura turística. PhD Tesis, University of Aveiro

European Commission [EC] (2006) Methodological work on measuring the sustainable development of tourism—Part 1: Technical report. Office for Official Publications of the European Communities, Luxembourg. http://bit.ly/1tfuUW0. Accessed 16 June 2013

European Environment Agency [EEA] (2005) EEA core set of indicators—guide (EEA technical report n.1/2005). Office for Official Publications of the European Communities, Luxembourg

Gawlik E, Kabaria H, Kaur S (2011) Predicting tourism trends with Google Insights. http://stanford.io/V11AWl. Accessed 1 Dec 2012

Gretzel U, Fesenmaier D, Leary J (2006) The transformation of consumer behaviour. In Buhalis D, Costa C, Ford F (eds) Tourism business frontiers: consumers, products and industry (pp. 9–18). Burlington: Butterworth-Heinemann/Elsevier

Gretzel U, Sigala M, Xiang Z, Koo C (2015) Smart tourism: foundations and developments. Electron Markets 2015(25):179–188

Gurău C (2008) Integrated online marketing communication: implementation and management. J Commun Manag 12(2):169–184

Hill MM, Hill A (2002) Investigação por Questionário, 2nd edn. Edições Sílabo, Lisboa

Höjer M, Wangel J (2015) Smart sustainable cities: definition and challenges. In: Hilty LM, Aebischer B (eds) ICT innovations for sustainability. Advances in intelligent systems and computing 310. Springer International Publishing, Swizerland
Hugony C, Cladera JR (2008) Indicadores para la Evaluación de las Ciudades Históricas. ACE-Arquitectura, Ciudad y Entorno III(8)
Instituto Nacional de Estatística [INE] (2012a) Anuário Estatístico da Região do Norte 2011. Instituto Nacional de Estatística, Lisboa
Instituto Nacional de Estatística (2012b) Anuário Estatístico da Região do Centro 2011. Instituto Nacional de Estatística, Lisboa
Instituto Nacional de Estatística (2012c) Anuário Estatístico da Região de Lisboa 2011. Instituto Nacional de Estatística, Lisboa
Instituto Nacional de Estatística (2012d) Anuário Estatístico da Região do Alentejo 2011. Instituto Nacional de Estatística, Lisboa
Instituto Nacional de Estatística (2012e) Anuário Estatístico da Região do Algarve 2011. Instituto Nacional de Estatística, Lisboa
Instituto Nacional de Estatística (2012f) Anuário Estatístico da Região da Madeira 2011. Instituto Nacional de Estatística, Lisboa
Instituto Nacional de Estatística (2012g) Anuário Estatístico da Região dos Açores. Instituto Nacional de Estatística, Lisboa
Jackman M, Naitram S (2015) Nowcasting tourist arrivals to Barbados. Just Google It! Tour Econ 21:1309–1313
Jackson S (2009) Cult of analytics: driving online marketing strategies using web analytics. Routledge, Oxford
Jones R, Zhang WV, Rey B, Jhala P, Stipp E (2008) Geographic intention and modification in web search. Int J Geogr Inf Sci 22(3):229–246
Kaushik A (2010) Web analytics 2.0: the art of online accountability & science of customer centricity. Wiley, Indianapolis
Leung D, Law R, Hoof H, Buhalis D (2013) Social media in tourism and hospitality: a literature review. J Travel Tour Market 30:3–22
Li X, Pan B, Law R, Huang X (2017) Forecasting tourism demand with composite search index. Tour Manag 59:57–66
Li J, Xu L, Tang L, Wang S, Li L (2018) Big data in tourism research: a literature review. Tour Manag 68:301–323
Milano R, Baggio R, Piatelli R (2011) The effects of online social media on tourism websites. ENTER 2011-18th International Conference on Information Technology and Travel & Tourism, 26–28 January 2011. Innsbruck, Austria
Monjardino I (2009) Indicadores de Sustentabilidade do Turismo nos Açores: o papel das opiniões e da atitude dos residentes face ao Turismo na Região. Artigo apresentado no 15º Congresso da APDR – Redes e Desenvolvimento Regional, Praia, Cabo Verde, 6 a 11 de julho
OECD (2003) OECD environmental indicators: development, measurement and use. Reference paper. OECD, Paris
OECD (2008) Handbook on constructing composite indicators: methodology and user guide. OECD-JRC. OECD Publishing, Paris
Pan B, Litvin S, Goldman H (2006) Real users, real trips, and real queries: an analysis of destination search on a search engine. Paper presented at the Annual Conference of Travel and Tourism Research Association (TTRA 2006). Ireland, Dublin, 16–18 June
Pan B, Wu D, Song H (2012) Forecasting hotel room demand using search engine data. J Hosp Tour Technol 3(3):196–210
Pérez VE, Blancas FJ, González M, Guerrero FM, Lozano M, Pérez F, Caballero RE (2009) Evaluación de la Sostenibilidad del Turismo Rural mediante indicadores sintéticos. Revista Investigación Operacional 30(1):40–51
Poon A (1993) Tourism technology and competitive strategies. CAB International, Oxford

Resolução do Conselho de Ministros n.º 24/2013 de 16 de abril. Diário da República n.º74/2013 – I Série. Presidência do Conselho de Ministros. Lisboa
Rheem C (2012) Empowering inspiration: the future of travel search. PhoCusWright Inc., Sherman. http://bit.ly/1qo6NNR. Accessed 30 June 2014
Ritchie JRB, Crouch GI (2003) The competitive destination: a sustainable tourism perspective. CABI Publishing, London
Saidi N, Scacciavillani F, Ali F (2010) Forecasting tourism in Dubai. Economic Note n.º 8, Dubai International Finance Centre. http://www.difc.ae/publications. Accessed 6 Aug 2011
Sanderson M, Kohler J (2004) Analyzing geographic queries. In the proceedings of SIGIR Workshop on Geographic Information Retrieval: The 27th Annual International ACM SIGIR Conference. Sheffield, UK, 25–29 July 2004
Shimshoni Y, Efron N, Matias Y (2009) On the predictability of search trends (draft). Google, Israel Labs. http://bit.ly/1zjL573. Accessed 15 Nov 2012
Smith E, White S (2011) What insights can Google Trends provide about tourism in specific destinations? ONS, UK. http://bit.ly/1o13x0r. Accessed 10 June 2012
SREA, DREM, ISTAC (2006) Sistema de Indicadores de Sustentabilidade do Turismo da Macaronésia 2000–2005. Direcção Regional de Estatística da Madeira, Funchal
StatCounter (2018a) StatCounter Global Stats: search engine. *StatCounter*. http://bit.ly/2EoV0Pu. Accessed 15 Jan 2018
StatCounter (2018b) StatCounter Global Stats: search engine. *StatCounter*. http://bit.ly/2BsKROE. Accessed 15 Jan 2018
Suhoy T (2009) Query indices and a 2008 downturn: Israeli data. Research Department, Bank of Israel. Discussion Paper No. 2009.06. http://bit.ly/Z29izI. Accessed 15 Jan 2012
Tussyadiah I, Fesenmaier DR (2009) Mediating tourist experiences: access to places via shared videos. Ann Tour Res 36:24–40
UNWTO (1996) What tourism managers need to know: a practical guide to the development and use of indicators of sustainable tourism. World Tourism Organization, Madrid
UNWTO (2004) Indicators of sustainable development for tourism destinations: a guidebook. World Tourism Organization, Madrid
UNWTO (2011) World Tourism Organization (UNWTO) Affiliate Members: AM-reports—technology in tourism, vol 1. UNWTO. http://bit.ly/ZtWjXd. Accessed 10 Aug 2014
UNWTO & ETC (2008) Handbook on E-marketing for tourism destinations. World Tourism Organization, Madrid
White V, McCrum G, Blackstock KL, Scott A (2006) Indicators and sustainable tourism: literature review. Unpublished report. The Macaulay Institute, Aberdeen
Xiang Z, Fesenmaier DR (2006) Assessing the initial step in the persuasion process: meta tags on destination marketing websites. Inf Technol Tour 8:91–104

Chapter 2
Developing Smart Tourism Destinations with the Internet of Things

Nicholas Wise and Hadi Heidari

Abstract The internet of things (IoT) aims to connect the objects of everyday life by embedding internet-connected devices within them and sharing their information online. Smart technology that exploits IoT data offers new opportunities for the travel and hospitality industry. The IoT enables easy access and interaction with a wide variety of information for contexts such as transportation, attractions, tours, shopping and hotels. IoT big data tourism applications will need to integrate social media, content marketing, and wearable IoT devices. After outlining conceptual understandings of the IoT and its potential for smart cities, this chapter provides practical foundations for destination organizers and stakeholders in this emerging smart tourism paradigm.

Keywords IoT · Smart cities · Smart destinations · Applications · Big data · IoT devices · Content marketing

2.1 Introduction

The internet of things (IoT) consists of everyday devices with embedded computing technologies that connect them to the internet. It allows new and more powerful applications to take advantage of new types of real-time data to deliver better services within the tourism sector and elsewhere. Creative uses of technology through devices like internet-connected and traffic-aware car navigation systems are already transforming our everyday activities, such as commuting. Our location-aware smartphone applications have also changed our overall spatial awareness, and give advice about how and what to consume when visiting a destination (Hedlund 2012; Vanolo 2014). The Internet now incorporates heterogeneously connected complex systems, such as wireless networks, sensors, actuators, and smart appliances (Li et al. 2015). Such

N. Wise (✉)
Liverpool John Moores University, Liverpool, UK
e-mail: N.A.Wise@ljmu.ac.uk

H. Heidari
University of Glasgow, Glasgow, UK

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_2

complex systems are further extending the virtual boundaries with physical entities and virtual components. The IoT will empower connected things with new capabilities in smart homes, smart cities, and smart wearable devices and clothes (Perera et al. 2015). The IoT will introduce new opportunities for the tourism industry by enabling easy access and interaction with a wide verity of information, integrating social media, content marketing, big data, and wearables. Consumers interacting with big data can transform the visitor experience, from knowledge production to knowledge exchange.

This chapter discusses the importance of the IoT for tourism to have transparent and seamless communications in complex systems involved in the development of the Big Data and Smart Cities. After outlining conceptual understandings of IoT in relation to tourism and destination management, this chapter will consider the usefulness of these points for the tourism industry. There is much that destination managers and organizers can exploit to develop better smart destinations. The internet marketing for tourism literature offers a starting point for destination organizers and a range of stakeholders to use in this emerging smart tourism paradigm. This chapter gives a conceptual overview of considerations, making it easier for consumers to find and experience visitor opportunities based on the IoT paradigm for tourism, and to understand this we will now discuss consider key literature on big data and smart connectivity before exploring the IoT for tourism.

2.2 Big Data and Smart Connectivity

A smart city uses data to supply information about people and for people to increase competitiveness, innovations and quality of life through enhanced connectivity (Albino et al. 2015). It also impacts tourists because they benefit from digital infrastructures, aggregated urban data, platform services, policies to enable processes and citizens in the city who supply information on the economy, community, culture and entertainment, movement and transport and urban places and spaces (Koo et al. 2017). Urban infrastructures include increasingly connectivity capacity, which has led to smart cities based on web-based platforms of communication and access (Lee et al. 2014) with the goal of improving citizens' quality of life (Kummithaa and Crutzen 2017). To improve living standards, both technology and human driven methods are helping to network places and enable participation to build knowledge societies (Kummithaa and Crutzen 2017). Then, to implement smart connectivity to enable big data collection, smart government, smart building, smart transport and smart utilities need equal investment to enable connections between government, businesses and citizens (and tourists) (Albino et al. 2015), as well as service, surveillance and distribution (Hancke et al. 2013). Whilst urban sustainability is often assessed based on social, economic and environmental impacts (Wise 2016), new technological advances target economic prosperity, ecological integrity and social equity, each based on connecting and informing tourists and consumers to enhance knowledge development (see Ericsson 2016; Gibbs et al. 2013; Neirotti et al. 2014). There is also potential for increasing social inequality because not all citizens will have easy access to necessary devices and internet access (see Calzada and Cobo 2015). The necessary communication infrastructures include "urban apps, big data, intelligent infrastructure, city sensors, urban dashboards, smart meters, smart buildings, and smart grids" (Luque-Ayala and Marvin 2015, p. 2107), useful for connecting businesses locals, and tourists through online participation (Afzalan et al. 2017; Hancke et al. 2013).

Smart tourism can build on notions of smart cities, not only because is it about urban quality of life, but it is also about enhancing the quality and experience of the destination based on value created, exchanged and consumed (Gretzel 2011). Online access and smart mobile devices are increasingly useful for business owners, managers, planners, locals and tourists. While many platforms share and communicate information, knowing how and when to communicate (and compute) data is important so that knowledge is transferred through appropriate communication channels. Smart understandings require copious amounts of data to create outcomes for people (Li et al. 2017), thus smart tourism needs to be part of the broader development and growth of smart cities. Moreover, digital connectivity can help locate visitors and position them in their new surroundings, helping a destination to target information at them (Borseková et al. 2017).

The criteria for maintaining competitiveness is regularly changing, so destinations need to adapt to technological changes. Going beyond the 6As of successful destinations (Attraction, Accessibility, Amenities, Availability, Activities and Ancillaries) (see Buhalis and Amaranggana 2014), the 6As should be framed based on interaction and real-time data (Brandt et al. 2017). Building on traditional forms of destination management, value creation requires accessible experiences to build and locate knowledge spatially (see Del Chippa and Baggio 2015; Del Vecchio et al. 2017; Zacatias et al. 2015). A conceptual framework for components is therefore needed to develop successful smart tourism destinations. For this, important components of smart data include information quality, source credibility, interactivity and accessibility (Yoo et al. 2017). A framework can be designed to inform travel decisions, with an emphasis on self-efficacy, guided by the user rather than by destination managers or tourism enterprises (Yoo et al. 2017). The nature of the data produced also shapes how users interact and consume in a destination, reinforcing the value of this data (Del Vecchio et al. 2017; Huang et al. 2017).

Sophisticated IoT sensors embedded in the physical things in the environment provide information and knowledge about many complex sensory systems. This information may be about providing more transportation choices or providing directions to the nearest hotel, dinning place or visitor attraction, for example. Generating and storing data using a common communication language is necessary for effective data integration. This requires an understanding of which data is available and how it can flow between different systems (Fig. 2.1). The process starts with the sensing devices securely communicating with an IoT platform. This data is transmitted between many devices and analytics to help deliver relevant information to applications that can more intelligently address industry and personal needs.

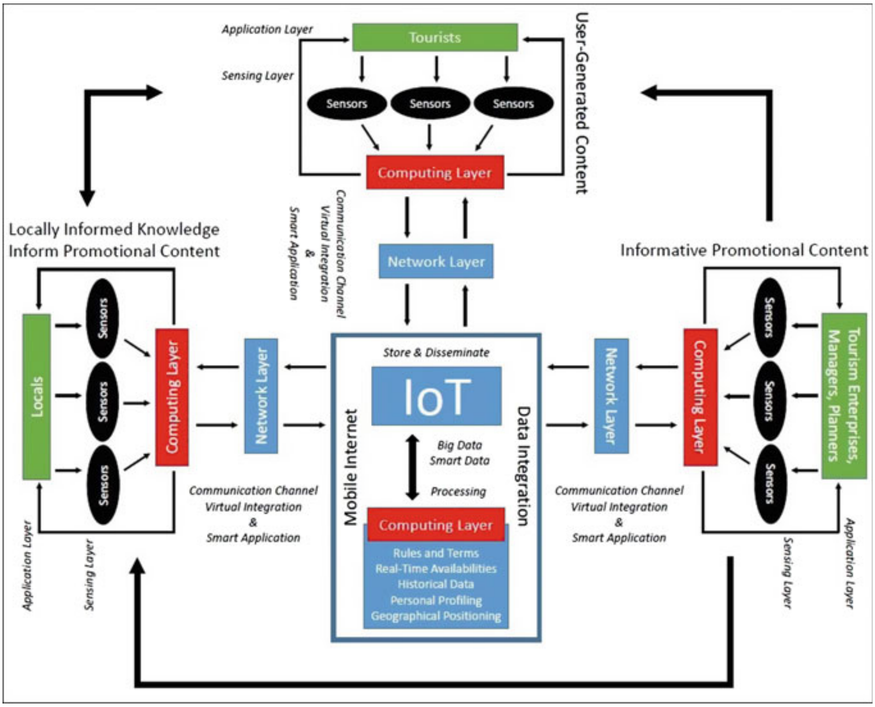


Fig. 2.1 Model of the data flows from the Internet of things to applications supporting tourism

The intelligent use of technology will not only improve existing services with real-time updates (like the localised real-time information provided by smartphone weather forecast apps) but will also create new personalised services to improve the overall travelling experience. Tourism is about consuming information at the appropriate time: providing the right service at the right time to the right person. Information delivered will be based on where people are (geographical positioning) rather than users needing to find relevant websites, allowing data to be delivered and authenticated instantly and in real-time (see Wise and Farzin 2018). Assistance can be provided using interactive cameras and embedded smart sensors around us that are already online and regularly updating, producing and processing data for us. Therefore, the IoT for tourism will change service by schedule to service by real time demand, potentially better matching the desires, demands and consumption needs of the tourist.

2.3 The Internet of Things for Tourism

Along with the popularity of people-centric sensing (Campbell et al. 2008a, b), IoT data fits into three categories: (1) personal sensing (2) social sensing and (3) public sensing. Whereas personal sensing is "focused on personal monitoring and archiving" the individual; social sensing is when "information is shared within social and special interest groups"; and public sensing occurs when "data is shared with everyone for the greater public good (such as entertainment or community action)" (Campbell et al. 2008b, p. 13). The emphasis across all three categories is "the ability to sense people and characteristics of their immediate surroundings, and the ability to sense data related to interactions between people and their surroundings" (Campbell et al. 2008a, p. 1).

Target data collected from each category contributes vital information in forming a network that will help to improve the travel experience. From personal food preferences to daily activities monitored by personal sensing, recognising individual differences will better enable IoT applications to suggest nearby offers best suited to personal consumption preferences. Examples include restaurants with favourite foods, locations that match a person's interests, or attractions that best suit a traveller's demographics (e.g., age, gender, nationality). The data sharing of personal activities recorded by sensing devices should be subject to ethical restrictions and privacy settings, including for modern wearables (Liang et al. 2017; Wen et al. 2016), connected cars (e.g. al-Khateeb et al. 2018) and wireless sensor networks (e.g. Alduais et al. 2017). Most importantly, connecting a device to the internet does not mean that its information is universally available and shared: the owner has some control over who can access the information. For example, a car navigation device will not broadcast the owner's location to anyone that is interested. Integrated cameras, GPS devices, microphones or accelerometers can communicate to either social or public sensing devices depending on the needs of an activity, such as community action, classes, entertainments, business, transportation or parks.

Mobile crowd sensing for smart cities can support efficient, safe and green mobility in urban environments (see Ganti et al. 2011; Pouryazdan et al. 2016). Given the ubiquity of mobile devices carried by people worldwide, social mobile crowd sensing through the IoT can allow tourists to know about popular events in a destination, provide interactive feedback with other tourists at different locations, reveal the best places to be at a certain time, local weather forecasts, and expected travel times throughout the day. Here crowd sourcing can inform people about whether to seek alternative routes, when best to arrive at attractions or restaurants, how to avoid unpleasant surprises when travelling, where to park, and which public transport solution would be best. Environmental sensors may also report air or noise pollutions levels. This enables tourists in unfamiliar places to make even better decisions than well informed locals might take. Box 2.1 illustrates the case of a tourist who travels to a new city and encounters an issue with their car.

Box 2.1 A Tourist Travels to a New City and Encounters an Issue with Their Car

A family travelling to new city by car for a holiday is alerted by the engine light that the car has a problem with its break line pressure that needs to be resolved quickly. Immediate attention from a mechanic is required, so the driver needs to find a nearby garage. She might also be notified by IoT devices about a nearby hospital and a local emergency telephone number in case there had been an accident. If she has to stop for a day while the car is fixed, her smartphone might recommend nearby hotels with vacancies that are suitable for a family and near to family-friendly attractions so that the unscheduled stop is not a disaster for the holiday.

Unknown to the tourist, businesses algorithms may access whatever information she has shared in public and use the decisions that she made about where to stay and where to visit so that they can develop better offerings and marketing strategies in the future, or target existing offerings at more suitable potential customers.

It is important and challenging to extract meaningful information from masses of raw data in an efficient way, such as by using event-linked networks (e.g. Sun et al. 2014), efficient maintenance and data management strategies (e.g. Zhuge and Sun 2010), and thorough data assessment and information organisation (e.g. Sun and Jara 2014). Linking the above context and understanding, the IoT will help local enterprises build awareness by informing relevant tourists about their locations, services and popularity (based on user-generated content provided by previous clients). Shops that are only known locally could increase their market-share through IoT and attract more customers. For instance, the small local business will get an opportunity to build a lasting online reputation from each customer, even if they only visit once, providing the customer needs are satisfactorily met. This may increase competition for high quality offerings within local communities, encouraging industries to sustain quality at a reasonable price. To attract tourism for economic growth, the environment needs to be greener and safer too. The preserving of natural and cultural heritage(s) are part of a place's identity, a source of attraction threatened by economic and/or social change (Hall 2015). The IoT can also help tourism indirectly through the systems that monitor environmental health issues, such as pollution levels.

The internet has long ago enhanced the ability to market destinations and encourage visits (Soteriades 2012), with all destinations today having an online presence. However, in an increasingly competitive tourism marketplace, tourists need fast, efficient and reliable information. This involves co-creation (Buonincontri and Micera 2016; Vicini et al. 2012) as destination managers and planners have similar information needs. Smart IoT tourism systems interact with tourists, enabling them to collectively engage and consume insightfully (Gretzel 2011). Geographically informed concentrations can create hotspots or paths, guided by local insight and business tactics to attract consumers (Hospers 2010). Insight and interest then become reinforced and supported by users, which is user-generated and/or user-guided, taking away from more traditional forms of marketing and destination planning (Cacho et al. 2016; Easton and Wise 2015).

2.4 Summary

Online content is useful for solving tourists' problems, and there are two main information sources (Buhalis and Amaranggana 2014): data supplied by the city and data from citizens/visitors. This alters the suppliers of data and information, as more traditional tourism information services need to change and adapt their approaches (see Li et al. 2017). From Fig. 2.1, local residents as data providers have dual roles. Locals can help confirm and inform promotional content as supplied by tourism enterprises (and business owners), and tourists who seek out local experiences can generate new data by their actions and preferences that can be intelligently analyzed in conjunction with other IoT data.

For each visitor, a pleasant journey involves minimising delays and unforeseen disappointments by knowing exact directions or times when an attraction is quiet. IoT applications have already started to allow travel to be more convenient and customised. By presenting more relevant, intelligent and customised information to the tourist, IoT-based services can support better decision making. The IoT can also help to improve the balance of the local economy, enabling local enterprises to compete for a larger market-share and learn how to improve their quality of service based on implicit feedback (e.g., whether users tend to select a service when presented it as part of a range of options) and user-generated content from consumers. With the proper integration of existing technologies, the future of IoT for the tourism industry suppliers, mangers and planners means better linking tourists based on local knowledge and informative promotional content.

References

Afzalan N, Sanchez TW, Evans-Cowley J (2017) Creating smarter cities: considerations for selecting online participatory tools. Cities 67(1):21–30

Albino V, Berardi U, Dangelico RM (2015) Smart cities: definitions, dimensions, performance, and initiatives. J Urban Technol 22(1):3–21

Alduais NAM, Abdullah J, Jamil A, Heidari H (2017) Performance evaluation of real-time multivariate data reduction models for adaptive-threshold in wireless sensor networks. IEEE Sens Lett 1(6):1–4

al-Khateeb H, Epiphaniou G, Reviczky A, Karadimas P, Heidari H (2018) Proactive threat detection for connected cars using recursive Bayesian estimation. IEEE Sens J 18(12):4822–4831

Borseková K, Vaňová A, Vitálišová K (2017) Smart specialization for smart spatial development: innovation strategies for building competitive advantages in tourism in Slovakia. Socio-Econ Plann Sci 58(1):39–50

Brandt T, Bendler J, Neumann D (2017) Social media analytics and value creation in urban tourism ecosystems. Inf Manage 54(6):703–713
Buhalis D, Amaranggana A (2014) Smart tourism destinations. In: Xiang Z, Tussyadiah I (eds) Information and communication technologies in tourism. Springer, Berlin, pp 553–564
Buonincontri P, Micera R (2016) The experience co-creation in smart tourism destinations: a multiple case analysis of European destinations. Inf Technol Tourism 16(3):285–315
Cacho A et al (2016) Social smart destination: a platform to analyze user-generated content in smart tourism destinations. In: Rocha A, Correia A, Adeli H, Reis L, Mendonça Teixeira M (eds) New advances in information systems and technologies. Advances in intelligent systems and computing. Springer, Berlin, 817–826
Calzada I, Cobo C (2015) Unplugging: deconstructing the smart city. J Urban Technol 22(1):23–43
Campbell AT, Eisenman SB, Lane ND, Miluzzo E, Peterson RA (2008a) People-centric urban sensing. In: Proceedings of the 2nd annual international workshop wireless internet. Retrieved from http://doi.acm.org/10.1145/1234161.1234179
Campbell AT, Lane, ND, Miluzzo E, Peterson RA, Lu H, Zheng X et al (2008b) The rise of people-centric sensing. IEEE Internet Comput 12(4):12–21
Del Chippa G, Baggio R (2015) Knowledge transfer in smart tourism destinations: analyzing the effects of a network structure. J Destination Mark Manage 4:145–150
Del Vecchio P, Mele G, Ndou V, Secundo G (2017) Creating value from social big data: implications for smart tourism destinations. Inf Process Manage. https://doi.org/10.1016/j.ipm.2017.10.006
Easton S, Wise N (2015) Online portrayals of volunteer tourism in Nepal. Worldwide Hospitality Tourism Themes 7(2):141–158
Ericsson (2016) Laying the foundations for a smart, sustainable city. *Ericsson White Paper*. Retrieved from https://www.ericsson.com/assets/local/publications/white-papers/wp-smart-cities.pdf
Ganti RK, Ye F, Lei H (2011) Mobile crowd sensing: current state and future challenges. IEEE Commun Mag 49(11):32–39
Gibbs D, Krueger R, MacLeod G (2013) Grappling with smart city politics in an era of market triumphalism. Urban Stud 50(11):2151–2157
Gretzel U (2011) Intelligent systems in tourism: a social science perspective. Ann Tourism Res 38(3):757–779
Hall CM (2015) Heritage, heritage tourism and climate change. J Heritage Tourism 11(1):1–9
Hancke GP, de Carvalho e Silva B, Hancke Jr GP (2013) The role of advanced sensing in smart cities. Sensors 13(1):393–425
Hedlund J (2012) Smart city 2020: technology and society in the modern city. Microsoft Services, Seattle
Hospers GJ (2010) Making sense of place: from cold to warm city marketing. J Place Manage Dev 3(3):182–193
Huang DC, Goo J, Nam K, Yoo CY (2017) Smart tourism technologies in travel planning: the role of exploration and exploitation. Inf Manag 54(6):757–770
Koo C, Park J, Lee J-N (2017) Smart tourism: traveler, business, and organizational perspectives. Inf Manag 54(6):683–686
Kummithaa RKR, Crutzen N (2017) How do we understand smart cities? An evolutionary perspective. Cities 67(1):43–52
Lee JH, Hancock MG, Hu M (2014) Towards an effective framework for building smart cities: lessons from Seoul and San Francisco. Technol Forecast Soc Chang 89(1):80–99
Li S, Xu LD, Zhao S (2015) The internet of things: a survey. Inf Syst Front 17(2):243–259
Li Y, Hu C, Huang C, Duan L (2017) The concept of smart tourism in the context of tourism information services. Tour Manag 58(5):293–300
Liang X, Heidari H, Dahiya R (2017) Wearable capacitive-based wrist-worn gesture sensing system. IEEE New Gener CAS. Retrieved from http://ieeexplore.ieee.org/document/8052299/
Luque-Ayala A, Marvin S (2015) Developing a critical understanding of smart urbanism. Urban Stud 52(12):2105–2116

Neirotti P, Marco AD, Cagliano AC, Mangano G, Scorrano F (2014) Current trends in smart city initiatives: some stylised facts. Cities 33(1):25–36
Perera C, Liu CH, Jayawardena S (2015) The emerging internet of things marketplace from an industrial perspective: a survey. IEEE Trans Emerg Top Comput 3(4):585–598
Pouryazdan M, Kantarci B, Soyata T, Song H (2016) Anchor-assisted and vote-based trustworthiness assurance in smart city crowd sensing. IEEE Access 4:529–541
Soteriades M (2012) Tourism destination marketing: approaches improving effectiveness and efficiency. J Hospitality Tourism Technol 3(2):107–120
Sun Y, Jara AJ (2014) An extensible and active semantic model of information organizing for the Internet of Things. Pers Ubiquit Comput 18(8):1821–1833
Sun Y, Yan H, Zhang J, Xia Y, Wang S, Bie R, Tian Y (2014) Organizing and querying the big sensing data with event-linked network in the Internet of Things. Int J Distrib Sens Netw 10(8):1–11
Vanolo A (2014) Smartmentality: the smart city as disciplinary strategy. Urban Stud 51(5):883–898
Vicini S, Bellini S, Sanna A (2012) How to co-create Internet of things-enabled services for smarter cities. In: SMART 2012: the first international conference on smart systems, devices and technologies. Retrieved from https://www.thinkmind.org/index.php?view=article&articleid=smart_2012_2_50_40077
Wen S, Heidari H, Vilouras A, Dahiya R (2016) A wearable fabric-based RFID skin temperature monitoring patch. In: Proceedings IEEE Sensors conference. Retrieved from http://ieeexplore.ieee.org/document/7808919/
Wise N (2016) Outlining triple bottom line contexts in urban tourism regeneration. Cities 53(1):30–34
Wise N, Farzin F (2018) See You in Iran" on Facebook: assessing "user-generated authenticity. In: Rickly J, Vidon E (eds) Authenticity & tourism: productive debates, creative discourses. Emerald, Bingley, UK, pp 33–52
Yoo CW, Goo J, Huang CD, Nam K, Woo M (2017) Improving travel decision support satisfaction with smart tourism technologies: a framework of tourist elaboration likelihood and self-efficacy. Technol Forecast Soc Chang 123:330–341
Zacatias F, Cuapa R, de Ita G, Torres D (2015) Smart tourism in 1-click. Procedia Comput Sci 56(1):447–452
Zhuge H, Sun Y (2010) The schema theory for semantic link network. Future Gener Comput Syst 26(3):408–420

Chapter 3
Big Data in Online Travel Agencies and Its Application Through Electronic Devices

Josep Ma Espinet

Abstract The aim of this chapter is to shed light on the present and future of big data in OTAs with regard to the particular electronic device where the website or app is displayed. Use of big data is strategic for OTAs as it could allow these companies to gain a competitive advantage and reduce their costs. The results from the empirical research carried out specifically for this chapter reveal that OTAs use big data extensively throughout the entire customer experience. Nevertheless, untapped potential remains which could be exploited to derive further competitive advantage. The main differences created by using different electronic devices is the quantity of information displayed due to the reduced size of the screens. Smartphones can provide OTAs another important difference through the use of GPS and highly accurate tracking technologies that enable these companies to obtain accurate information about what their customers do during a stay in a destination, so that these companies can offer a more customized service. Finally, OTAs should consider big data as a mindset which affects the whole company and its organizational structure, and not only as information and its associated technology.

Keywords OTAs · Electronic devices · Big data · Customer experience · Smartphones

3.1 Introduction

One Sunday my wife and I were comfortably sitting at home trying to find accommodation for our summer holidays on some online travel agencies (OTA). One of us was looking for information on the laptop and the other was doing the same on a smartphone. When we compared the results, we realized that although we had been searching in the same OTA and we had introduced the same parameters (city, dates

J. M. Espinet (✉)
Universitat de Girona, Girona, Spain
e-mail: jmaespinet@gmail.com

J. M. Espinet
Universitat de Mediterrani, Barcelona, Spain

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_3

and filters), we had obtained different hotels and viewed different information. Could that be possible?

OTAs are a common way of booking accommodation, flights and other types of tourism services. The most important OTAs (e.g., Booking, Expedia, and TripAdvisor) offer services and are available around the world. In fact, the competitive advantage of OTAs is that they provide a useful service to suppliers, which are sometimes an important part of their sales and can have a dominant position in the marketing of hotels (Murphy et al. 2016), as well as to customers, for whom it is a simple and easy way to make a booking at any time. Furthermore, technological advances have allowed consumers to access OTAs using a variety of electronic devices such as desktops, laptops, tablets or smartphones. As these enterprises sell millions of rooms per year, their transactions generate a vast quantity of information that can be very useful for big data development and exploitation. In the business model of OTAs, big data acquires special relevance by making it possible to trace all the movements of users and providing useful information at any stage of the customer journey map.

The use of different types of electronic devices is widespread. Ditrendia (2017), in research conducted in 2016, showed that mobiles are the electronic devices most used to access the Internet, followed by laptops, desktops and tablets. Among people who used electronic devices, 57% did it on more than one type once a day and 21% used more than one device at the same time. The same research revealed that the average user spent 170 min/day on their smartphone and more than 50% of global Google searches were conducted from smartphones. Hence, smartphones, which are always within easy reach of the user, are expected to perform the same functions as larger devices, which require adaptations of software, interface and contents.

The aim of this chapter is to analyze how OTAs are currently using big data and to anticipate how the industry might be able to exploit it in the future, especially in terms of the different types of electronic devices used to access their services. More specifically, the research questions are:

- How are OTAs using big data as a whole and adapting their services to best suit different electronic devices?
- How can these strategies benefit OTAs and customers?

These questions will be answered by considering the big data challenges described by Sivarajah et al. (2017, p. 265), who reviewed relevant research on the topic and undertook extensive research of their own.

To sum up, this chapter presents:

- A review of the most relevant literature
- Empirical evidence showing how OTAs use big data depending on the electronic device is being used and how they face up their challenges
- Ways in which big data could potentially be exploited to benefit OTAs and create customer value.

3.2 Theoretical Framework

Analyzed from the view of practitioners and academics, big data is more than technology, software or the integration of different sources of information (Gandomi and Haider 2015). Big data does not mean immediate success, and its development is not cheap, easy or fast. Therefore, prior to making large investments in it, companies should assess whether big data is likely to offer a return and should understand how it can be optimally used. Baldwin (2015) points out that nearly 80% of companies fail to fully integrate their data into their company practices and 65% consider their data management practices to be weak. The same study reported that 67% of surveyed companies had no "well-defined criteria to measure the success" of big data investments. In order to tackle this situation, it is proposed suboptimal developments that do not produce worse forecasting performance over time (Nikolopoulos and Petropoulos 2017) or step-by-step development, both of which allow companies to progressively learn from the experience.

Big data challenges, according to Sivarajah et al. (2017, p. 265), can be classified into three categories: data (volume, velocity, variety, variability, veracity, visualization and value), process (data acquisition and warehousing, data mining and cleansing, data aggregation and integration, analysis and modelling, and data interpretation) and management (privacy, security, data governance, data and information sharing, cost/operational expenditures, data ownership). The same authors describe the types of big data analytical methods as descriptive, inquisitive, predictive, prescriptive and pre-emptive depending on the stage of exploitation-information, insight, decision and action (p. 266).

Raguseo (2018) empirically analyzes the benefits and the risks of the adoption of big data technologies in companies by giving additional consideration to their size and their industrial sector. The author concludes that portal content is the most frequently used big data source, that visual analytics software, scripting languages and in-memory are the technologies most frequently adopted, and that privacy and security are the most frequently mentioned risks. Moreover, she indicates that, regardless of the sector, companies apply the same big data technologies; the differences are in the size of the companies. Côrte-Real et al. (2017) point out that big data analytics can create organizational agility through knowledge management, which can in turn lead to competitive advantage.

The impact of big data on competitive advantage is explained in Erevelles et al. (2016), and Ur Rehman et al. (2016) propose a win-win business model for value creation through big data and present a map of potential application areas on a business model canvas where mobile devices are the most common channels. The different aims of this model include lowering costs, enhancing the trust between customers and enterprises, preserving privacy of customers, and enabling secure data sharing.

Some of the research reviewing big data is more conceptual and some is more technical. Yaqoob et al. (2016) explain the past, present and future of big data; Sivarajah et al. (2017) define challenges and analytical methods; Erevelles et al.

(2016) point out its impact on competitive advantage and value creation; Kambatla et al. (2014) present an overview of the state of the art in big data analytics; Stieglitz et al. (2018) analyze the importance of social media analytics on big data; and Hashem et al. (2015) focus on cloud computing.

From the point of view of tourism research, extensive literature on big data can be found in Li et al. (2018). These authors classify big data in tourism research according to the data source—users, devices and operations—explaining the data types in each case. OTAs are an important source of data in studies in tourism, mainly through the use of online reviews. Data was obtained mainly from TripAdvisor, the most popular tourism social media and now an important OTA (Guo et al. 2017). Other studies collect data from Ctrip, Expedia, Booking, Qunar and Yelp. Xiang et al. (2017) provide an interesting overview by comparing reviews from TripAdvisor, Expedia and Yelp, and Marine-Roig and Clavé (2015) use the information of TripAdvisor and Venere to highlight the usefulness of big data analytics to support smart destinations. From the point of view of marketing, the visual impact is convenient (Cheng and Edwards 2015) and Google Analytics is a useful platform (Plaza 2011). Nevertheless, the specific strategies of big data in OTAs have not been analyzed in detail.

With regard to electronic devices, Kim and Law (2015) review the literature on tourism and hospitality marketing on smartphones and Murphy et al. (2016) is an important reference regarding the relationship between multiple devices and information sources used in hotel booking processes. Consultation via mobile phones is faster and more streamlined than on desktop or laptop computers so it is necessary for companies to adapt to new requirements. Smartphone applications provide access to location-based information, relevant to the immediate surroundings of tourists, and to variable content, which is timely and updated; flexibility in terms of delivering text, video or images; and interactive annotations which are integrated with map-based services and additional information (Yovcheva et al. 2012). Current technologies such as location-based services have the ability to easily capture spatial information of the user through mobiles and tablets providing data of great value (Ayscue et al. 2016). Hardy et al. (2017) tracked the movement of 472 tourists in real time via integrated surveys and GPS, discuss the implications of different approaches for tracking tourists, and provide innovative methods to use in smartphones.

This chapter offers a new perspective on the analysis of big data, analyzing the entire process followed by a user who is booking through an OTA and using different types of devices. The challenges identified by Sivarajah et al. (2017) are the main framework.

3.3 Methodology

To support comments, proposals and conclusions, empirical research[1] was carried out. The research undertaken was mostly qualitative and based on observation and

[1] Author has documentation of all the empirical research done.

Table 3.1 Selected OTAs and meta-search engines

7 ideas	Edreams	Hotelscombined (*)	Prestigia
Agoda	Expedia	HRS	Priceline
Amoma	Getaroom	In.via	Skoosh
Atrapalo	Hotel.info	Kayak (*)	TripAdvisor (*)
Booking	Hoteles	Lasminute	Trivago (*)
Centraldereservas	Hotelius	Laterooms	Venere
Ctrip	Hotelopia	Logitravel	Viajes El Corte Inglés
Destinia	Hotelsclick	Momondo (*)	*(*) Meta-search engines*

the extensive analysis of the main OTAs. The empirical strategies of big data were analyzed mainly in the first steps of the booking process, that is, at the point of first contact between the user and the OTA. More specifically, the analysis focuses on the registration requirements, the type of filters offered to select accommodation, and all the information displayed regarding the hotels that result from the search.

The information was obtained from the websites or apps of the OTAs through different types of electronic devices. In order to select the OTAs that were to be analyzed for this research, the words '*Hotel Booking*' and '*Reservar Hotel*' were entered as search terms on www.google.com, on 12 September 2017. The OTAs selected were those that appeared on the first three pages. Finally, 31 webpages from around the world were selected, corresponding to 26 OTAs and 5 meta-search engines—although these are not OTAs, they offer a similar service and were considered—(Table 3.1). This sample can be considered significant as this market has a high concentration of companies.

The most important OTAs around the world and in Spain—Booking, Edreams, Expedia, Logitravel, TripAdvisor and Trivago—were analyzed extensively, searching for accommodation in Barcelona from 15–17 May 2018. The data was gathered during the first week of January 2018 using four types of electronic devices at the same time—desktop (19′), laptop (15.6′), tablet (9.7′) and smartphone (4.7′)—in both websites and app format. Although websites and apps serve the same function, differences do exit. Apps are software applications that consumers must download onto a device, occupying space. They allow OTAs to offer more personalized service although some include technical developments that not all OTAs support.

3.4 Results

As mentioned in the introduction, OTAs offer a wide range of services and are essential for travelers when they compare and consider options (Verma et al. 2012) due in part to the easy comparisons, user-friendliness and mobile-friendly content (Murphy et al. 2016). Figure 3.1 shows the process followed by a user. The first contact users have with OTAs is when they access the website or app. From that point on the users' online activity can be traced, permitting OTAs to begin to gather and

Fig. 3.1 Process followed by a user when accessing an OTA webpage

exploit the potential of the data. This section explains the results from the empirical research.

3.4.1 OTA Webpage

When users access the webpage of an OTA, the first message they usually receive asks whether they accept the page's cookies policy (Fig. 3.2). In case of acceptance, OTA can log all subsequent actions from that IP address.

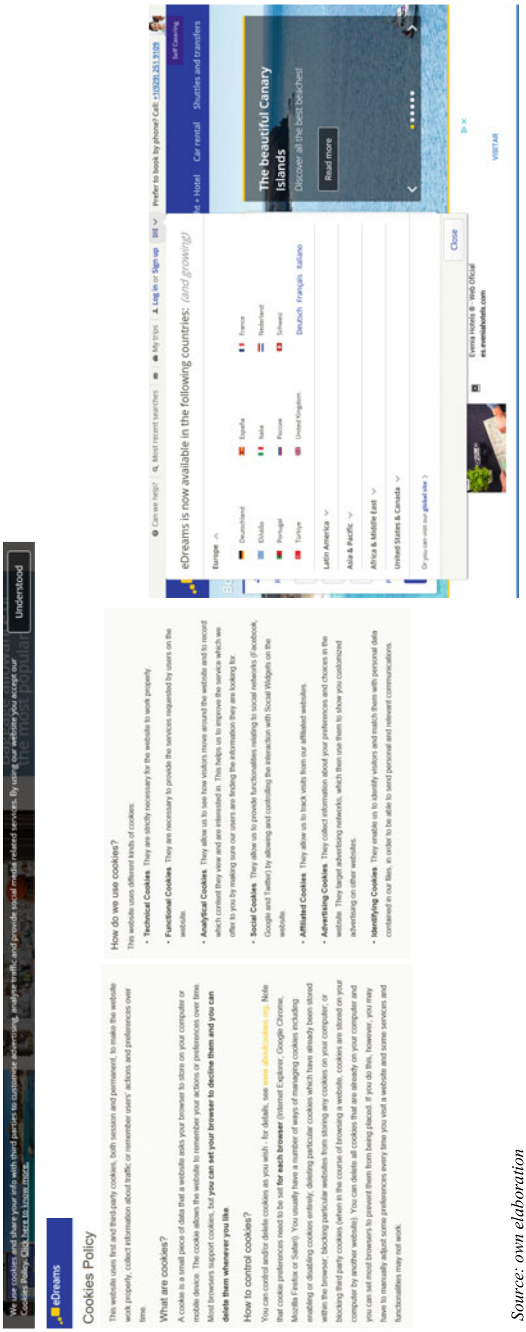
Frequently, in the following step, users either select the country where they are based or the language in which they want the webpage to be displayed. In both cases big data can be applied to develop specific strategies to attract customers.

3.4.2 Registration

Registration (Fig. 3.3) is optional and can provide benefits to OTAs and customers. Empirical analysis shows that not all OTAs (Table 3.2) offered the option of registration (28 out of 31 OTAs analyzed offered the possibility of registering via the web and 22 via an app). Of those OTAs that offered the option, the majority asked for 3 fields of which 1 was compulsory. As expected, no variation in the number of compulsory fields was observed regardless of the electronic device used. The reduced number of fields can be explained by a desire to facilitate the process of registration.

The most frequently requested compulsory field is an e-mail address, given that it is the easiest way to contact users, followed by the name and surname field. Although not always obligatory, 100% of the OTAs requested an e-mail address, 71% a Facebook account, 57% the name and surname, 43% a Google account, and 14% a mobile phone number. Other information requested included address, city, country, post code (each of these items were requested by approximately 7% of OTAs) and passport country, currency, preferred language (each of these items was requested by approximately 4% of OTAs).

Facebook profiles are among the most useful items for OTAs. Apart from being a simple way of registering, they also allow OTAs to access personal information that allows them to take advantage of big data. Nowadays, social media are very important sources of data that can be used in different ways. Khan et al. (2014) summarize the data generated by social media. Exploiting it requires a process based on discovery, collection and preparation. As happens in all data collection processes,



Source: own elaboration

Fig. 3.2 eDreams. Cookies policy and country of the user. *Source* Own elaboration

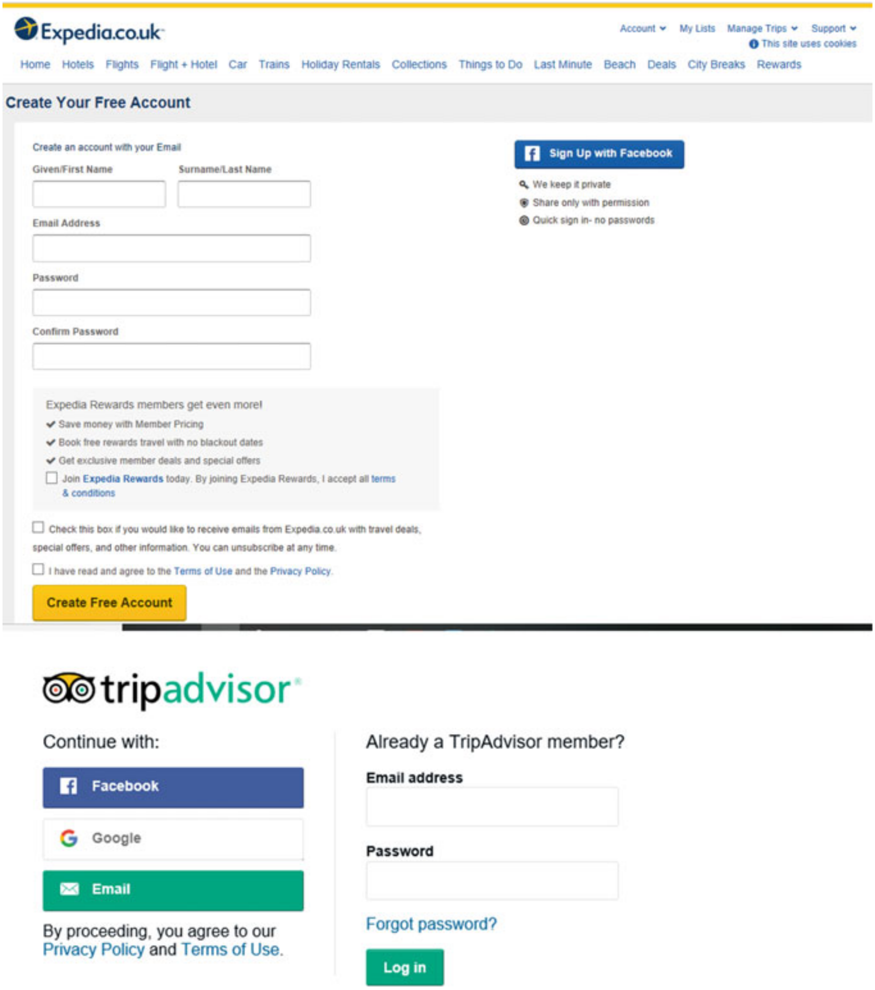


Fig. 3.3 Examples of fields requested to register

it is important to filter out irrelevant information and big data can provide algorithms that can partially solve the problems generated by this irrelevant information (Stieglitz et al. 2018). Facebook is being used as a marketing tool by destination management organizations (see the case of Italian in Mariani et al. 2016).

To sum up, registration is a simple process during which no difficult or confidential information is needed. In fact, it has been suggested that companies would reduce the amount of unimportant/irrelevant data gathered if they make better use of big data (Erevelles et al. 2016; Oracle 2013). The type of information requested can be used to contact the customer (e-mail address, Facebook account, mobile number)

Table 3.2 Registration fields in OTAs (December 2017)

	Web	App
Compulsory fields	**28**	**22**
1 field	20	15
2 fields	6	4
3 or more fields	2	3
Compulsory and optional fields		
1 field	1	3
2 fields	7	7
3 fields	14	9
4 fields	4	1
5 or more fields	2	2

and most fields can be exploited by big data (Facebook, location, address, city, post code, passport country, currency and language).

3.4.3 Filters

To help customers select accommodation, OTAs offer a variety of filters which narrow the search results, thereby facilitating the decision. At the time of research, variations in these filters depended on the type of device used, whether the format was web or app, and whether the user was registered or not.

From the analysis of the 31 OTAs on the previously mentioned types of electronic devices, more than 50 possible types of filters were observed, although some of them were similar. This is due to the growing number of possibilities that new platforms, interfaces and software provide OTAs. In fact, OTAs can parametrize websites and apps and easily and quickly create, change or delete filters as necessary. This is important as it makes the booking process easy, thereby facilitating customer decisions and encouraging brand loyalty. All these filters can be divided into two categories: those related to accommodation (i.e. attributes, facilities and characteristics), and those related to the area surrounding the accommodation (i.e. the distance to places of interest) that cannot be improved by the establishment. For example, after a stay, Booking sends a survey to the user that includes questions about the accommodation, the location and the area as a destination.

On average, OTAs offered 11.8 filters, with most offering between 5 and 14, although some of the bigger agencies such as Booking, Edreams and Trivago offered more than 20. As expected, the more reduced the screen space is, the fewer the available options. (Desktop formats show an average of 12.5 filters and smartphones formats 10.4.) The main differences were due to the type of electronic device used, while other factors (web vs. app, registered vs. non-registered) produced minimal

Table 3.3 Number of filters in webpages depending on the electronic device used

Number of filters	Desktop	Laptop	Tablet	Smartphone
Fewer than 5	0	0	0	0
5–9	8	8	9	14
10–14	17	19	18	14
15–19	3	1	1	2
From 20	3	3	3	1
TOTAL	31	31	31	31
Average	**12.5**	**12.3**	**11.9**	**10.4**

variation. These results confirm the conclusion of Murphy et al. (2016): the characteristics of tablets and smartphones result in fewer results per screen and make it more challenging for users to input data.

From the consumers' point of view, the more choices they have, the more complex and time consuming the process becomes. That is why some OTAs differentiate between 'top filters' (the most used) and 'extra filters'. Furthermore, some OTAs offer customers the opportunity to create their own filter and other OTAs personalize filters by identifying the most common filters previously used by a particular user or IP address. All these possibilities facilitate requests and the booking process (Table 3.3).

The first filters are usually the destination, the check-in and check-out dates, the number of adults and children, and the number of rooms. This order allows OTAs to make a more efficient use of software and platforms. At that point a great variety of filters are displayed. The most common filters are the star rating, which is usually the official rating awarded in that particular country (99%), budget (97%), facilities (86%), the name of the hotel (64%), guest rating (64%), type of property (58%) and location (54%). Other filters include the type of hotel (42%), the hotel chain (40%), whether the accommodation is offered at a promotional price (34%), distances from certain places (32%), places of interest (24%), type of rooms (21%), type of board (26%), room services (21%), top filters (17%), payment options (14%), free cancellation (13%), accessibility (12%), reason for travelling (12%), recommended filters (11%), proximity to public transport (11%), and TripAdvisor rating (11%). Figures 3.4, 3.5 and 3.6 show the filters offered by Booking, Trivago and TripAdvisor on laptops and smartphones.

Guo et al. (2017, p. 473) propose 19 controllable dimensions for hotel-customer interactions that could be very useful for hoteliers and also for OTAs. These dimensions are check-in and check-out, resort facilities, communication, homeliness, bathroom, room experience, events management, car parking, style and decoration, guest facilities in hotel, location in building (e.g. first floor), breakfast, value for money, price, staff service, room size, apartment, dining and accommodation for pets. On the other hand, the same authors identify three partially controlled dimensions—trans-

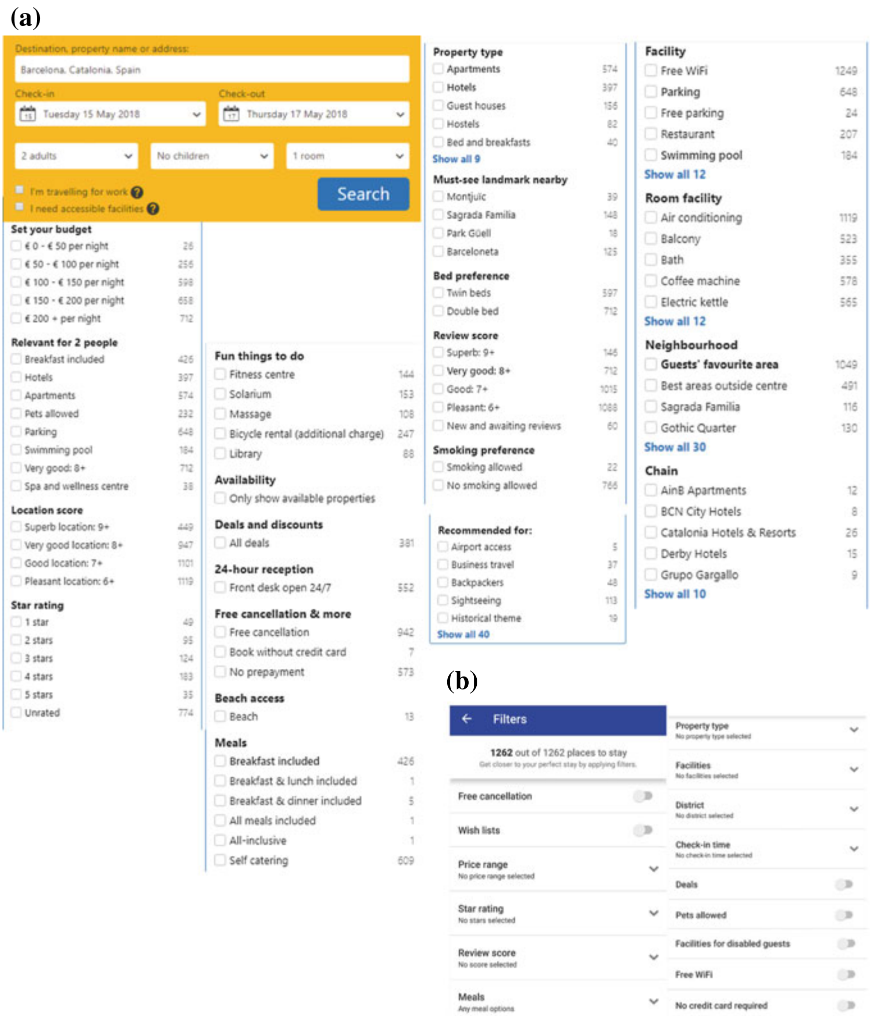


Fig. 3.4 **a** Laptop filters from www.booking.com (data obtained 16 January 2018). **b** Smartphone filters from www.booking.com (data obtained 16 January 2018)

port, location and visitor suitability—and three uncontrolled dimensions—weather, natural beauty and nightlife.

To sum up, the use of filters can be exploited by big data to facilitate user selection, benefitting, again, OTAs and customers.

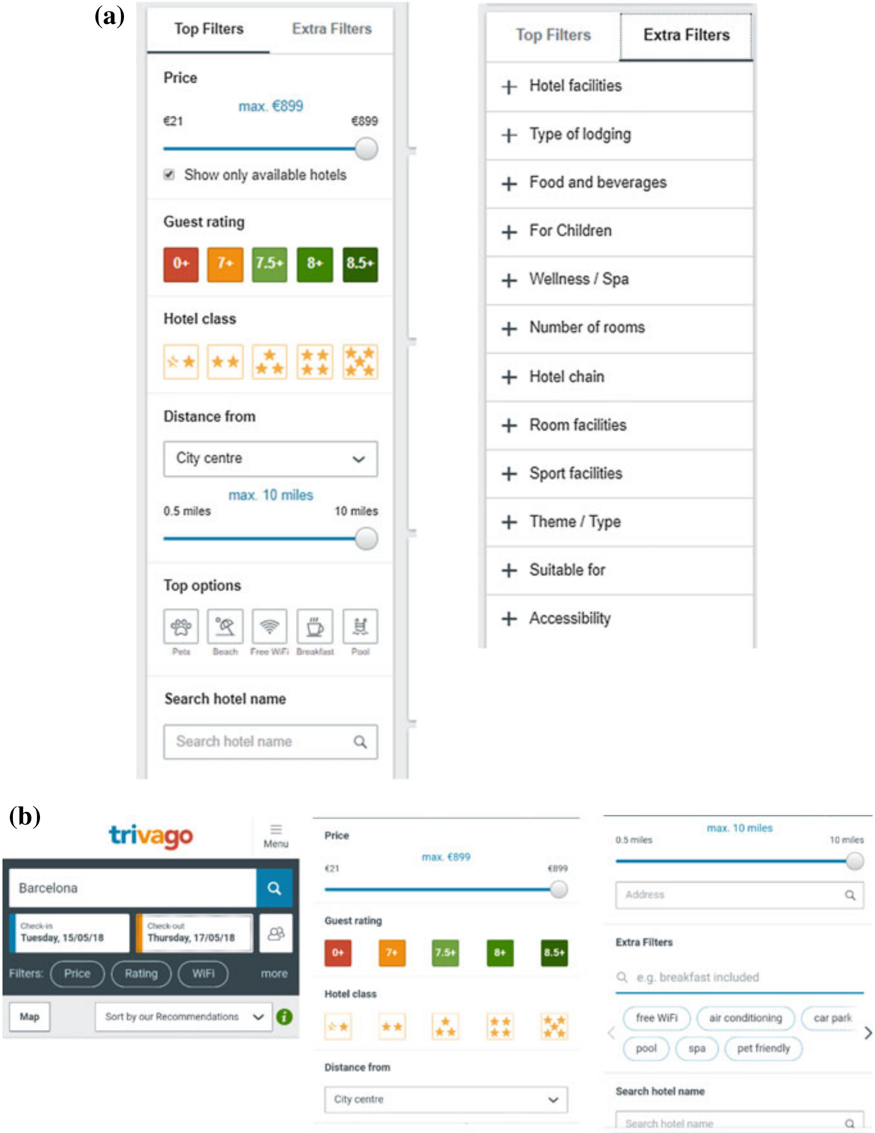


Fig. 3.5 **a** Laptop filters from www.trivago.com (data obtained 16 January 2018). **b** Smartphone filters from www.trivago.com (data obtained 16 January 2018)

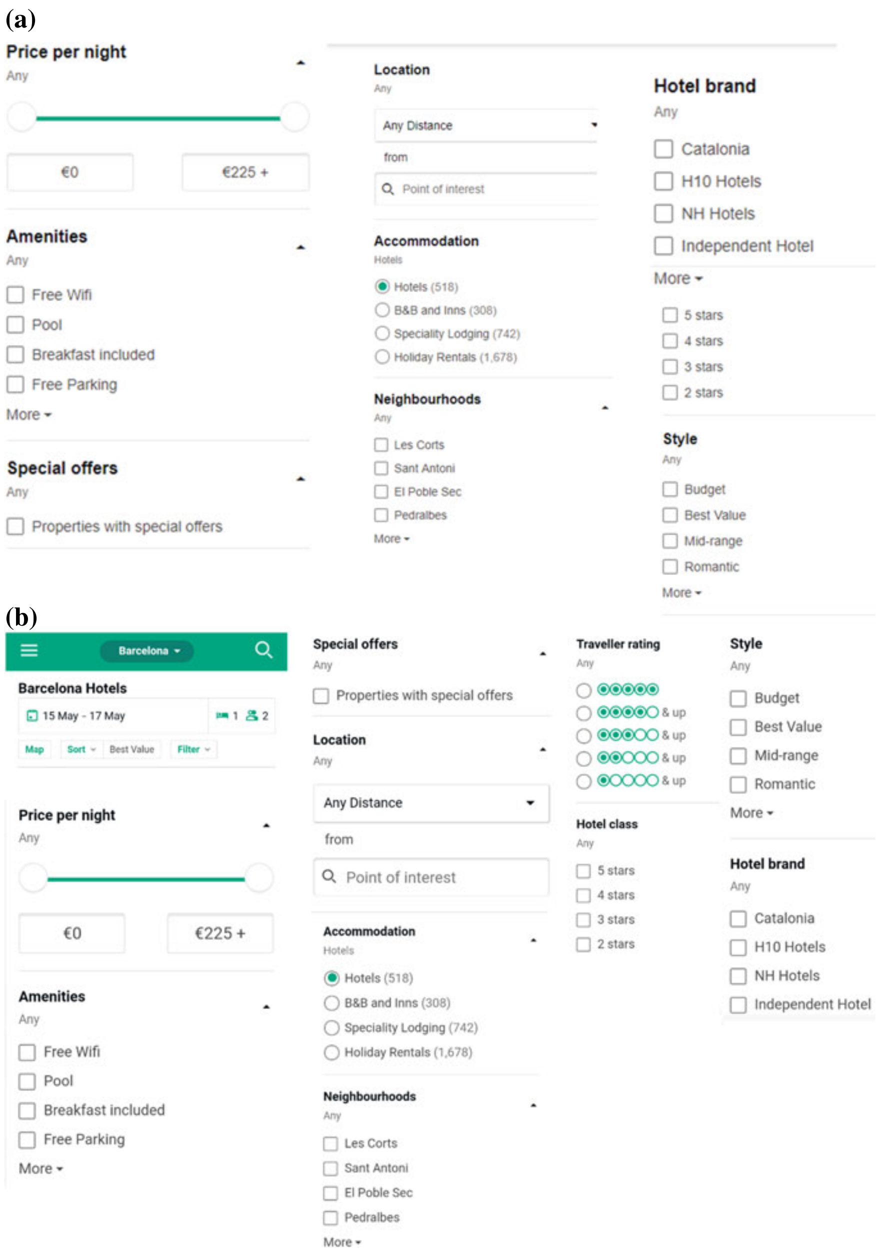


Fig. 3.6 **a** Laptop filters from www.tripadvisor.com (data obtained 16 January 2018). **b** Smartphone filters from www.tripadvisor.com (data obtained 16 January 2018)

3.4.4 Hotels Offered in the Search Results

After applying the filters, the website/app displays a list of hotels. The results reveal that some OTAs offer different hotels or the same hotels but in a different order depending on the format (see Table 3.4). The price, however, is always the same, although, in fact, it could be different. This could be a tactical decision or occasionally the result of a mistake. The six OTAs analyzed display a different number of hotels. Considering the first three screens, the average number of hotels displayed varies between 5.75 in Logitravel and 10.5 in Expedia (in the middle, Booking and TripAdvisor with 7.25, Edreams 7.5 and Trivago 9). On the other hand, the number of hotels displayed on the first three screens is different depending on the electronic device used (7.3 desktop, 6.3 laptop, 9.8 IPAD and 8 smartphone). Hence, visualization is an important decision that can affect user perception.

The information presented about each hotel is different depending on the format used (Figs. 3.7 and 3.8). The analysis of the 6 OTAs reveals that the information about each hotel that is always displayed is the price and the customer evaluation, which strongly influence customers as they decide to book a particular hotel or not. The number of comments (92%), the official category (83%), the location (62%), and the services (58%) are also frequently displayed. Then, by clicking on a specific accommodation, users are given extensive information including detailed opinions of customers, which can be seen as a bid to be transparent and to build customer trust. Again, the information displayed varies depending on the format.

Regarding the choice of destination, some OTAs indicate the percentage of reservations (level of occupancy) in a city during the selected dates. On occasions when occupancy is particularly high, alternative dates are offered along with the corresponding rates of occupancy so that customers can select dates when there will be a greater range of accommodations available and, presumably, at more affordable prices.

This empirical research also reveals that, due to the transparency of the market, it is evident that OTAs analyze their competitors, and consequently, the type of information displayed by competitors tends to be similar.

To sum up, OTAs can use big data in multiple ways when displaying hotels: first, to set prices and establish the order in which the hotels they offer are shown; second, in the information facilitated about each hotel; and third, in other information about the destination that can be useful for users. Moreover, the analysis of competitors can be exploited by big data tools.

3.4.5 Booking Process

Until this point, only data related to users' preferences in relation to the accommodation and destination under consideration has been collected. Much more specific and useful information can be obtained during the booking process by asking for

Table 3.4 Hotels offered in Barcelona from 15 to 17 May 2018 depending on the electronic device and after applying filters (information searched on 12 September 2017)

OTA	Position	Desktop	Laptop	IPAD	Smartphone
Booking	1	Hotel Lleó	Hotel Lleó	The Lonely Chimney Apartments	Apartments Casanova BCN
Booking	2	Travelodge Barcelona Poblenou	Travelodge Barcelona Poblenou	Hotel Vilamarí	Iberostar Paseo de Gracia
Booking	3	Hotel Medium Monegal	Hotel Medium Monegal	Apartments Casanova BCN	NH Barcelona Barri Gotic
Booking	4	Hotel Lloret Ramblas	Casagrand Luxury Suites	NH Barcelona Centro	Uma Suites Luxury Midtown
Booking	5	Ciutat Vella	Hotel Lloret Ramblas	Royal Ramblas	NH Collection Barcelona Podium
Expedia	1	Hotel Arts Barcelona	Hotel Arts Barcelona	Hotel Arts Barcelona	Hotel Arts Barcelona
Expedia	2	Arc la Rambla	Hotel Barcelona Universal	Barceló Raval	Barceló Raval
Expedia	3	Barceló Raval	Eurohotel Barcelona Gran Fira	Arc la Rambla	Arc la Rambla
Expedia	4	Hotel Porta Fira	Hostalbla	Hotel Barcelona Universal	Hotel Barcelona Universal
Expedia	5	EuroPark	Hotel REC Barcelona	Hotel Porta Fira	Hotel Porta Fira
Trivago	1	NH Collection Barcelona Podium	NH Collection Barcelona Podium	NH Collection Barcelona Podium	NH Collection Barcelona Podium
Trivago	2	Barcelona Princess	Barcelona Princess	Barcelona Princess	Barcelona Princess
Trivago	3	NH Barcelona Barri Gotic	NH Barcelona Barri Gotic	NH Barcelona Barri Gotic	NH Barcelona Barri Gotic
Trivago	4	Chic&basic Born	Chic&basic Born	Chic&basic Born	Chic&basic Born
Trivago	5	Mariano Cubi	Mariano Cubi	Mariano Cubi	Mariano Cubi

(continued)

Table 3.4 (continued)

OTA	Position	Desktop	Laptop	IPAD	Smartphone
Tripadvisor	1	SallesHotel Pere IV	Hotel SB Icaria Barcelona	Hotel El Avenida Palace	ABaC Barcelona
Tripadvisor	2	Room Mate Carla	Room Mate Emma	Renaissance Barcelona Hotel	H10 Port Vell
Tripadvisor	3	Room Mate Pau	SallesHotel Pere IV	Hotel Acta Atrium Palace	EuroPark Hotel
Tripadvisor	4	B-Hotel	Hotel Acta Atrium Palace	Room Mate Pau	Catalonia Catedral
Tripadvisor	5	Hotel Grums Barcelona	Room Mate Pau	Pol & Grace Hotel	Hotel Villa Emilia
Logitravel	1	Generator Hostel Barcelona	Generator Hostel Barcelona	Generator Hostel Barcelona	Generator Hostel Barcelona
Logitravel	2	Catalonia Park Güell	Catalonia Park Güell	Catalonia Park Güell	Catalonia Park Güell
Logitravel	3	Chic&basic Ramblas	Chic&basic Ramblas	Chic&basic Ramblas	Chic&basic Ramblas
Logitravel	4	Vincci Maritimo	Vincci Maritimo	Vincci Maritimo	Vincci Maritimo
Logitravel	5	Rialto		Rialto	Rialto

customer preferences. At this stage, customers can be asked which type of bed they prefer, whether they are interested in renting a car, what their estimated arrival time is, and whether they have any special requests. Again, this information can differ from OTA to OTA and from format to format.

When analyzed properly, this information is valuable in that it allows OTAs to offer services that more closely match customer preferences. It is also shared with the establishment in question, allowing them improve and provide greater customer satisfaction in the future.

3.4.6 Customer Experience

The period between a customer making a booking and arriving at the hotel can be a chance for OTAs to offer information about the destination or to cross-sell and up-sell. Again, many OTAs use big data to send customers reminders and to facilitate information they may need, which gives OTAs a competitive advantage

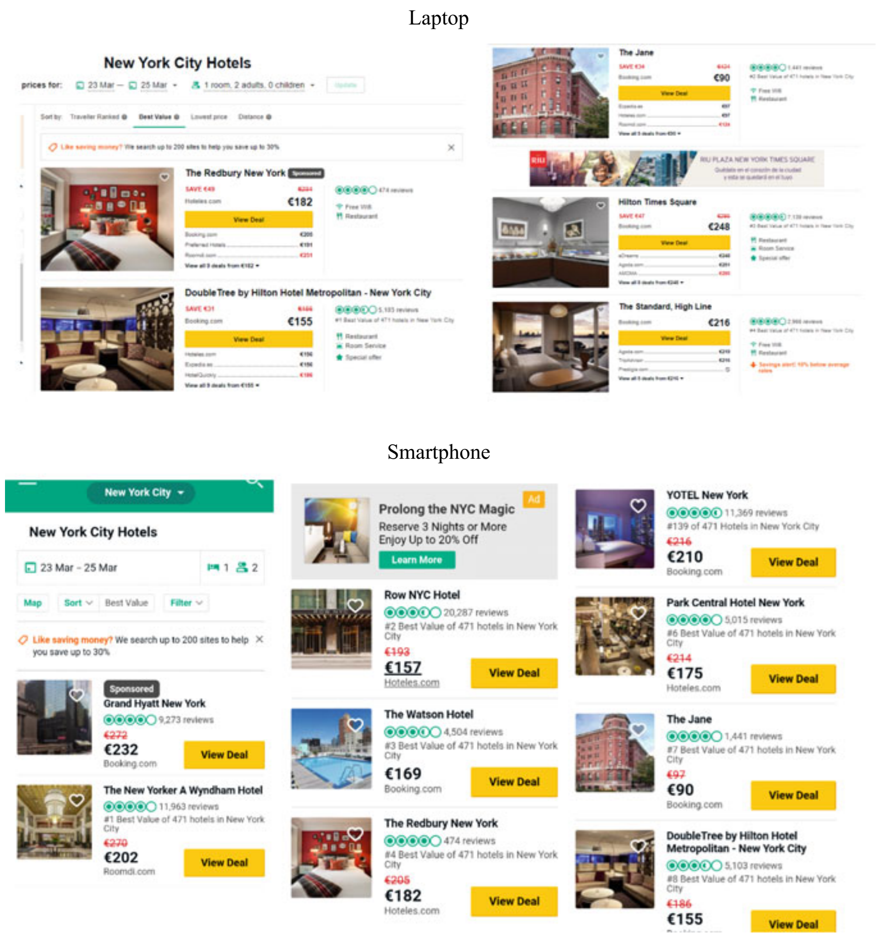


Fig. 3.7 Hotels displayed on www.tripadvisor.com in smartphones or on laptops at the same time (information searched on 26 January 2018)

over traditional travel agencies. Furthermore, some OTAs follow the customers' experience from the moment they check in at the hotel. For example, they might ask how satisfied a customer is with the check-in experience, a question which is most commonly posed via smartphone and answered using icons. OTAs could pass on this information when contacting the establishment in question.

Finally, most OTAs send a questionnaire to users after their stay to evaluate their level of satisfaction. This is very important because it gives OTAs a complete set of information about the user experience, permitting them to develop new strategies that benefit all parties (OTAs, hotel and future customers). Miah et al. (2017) present a method to extract, rank, locate and identify meaningful tourist information. The

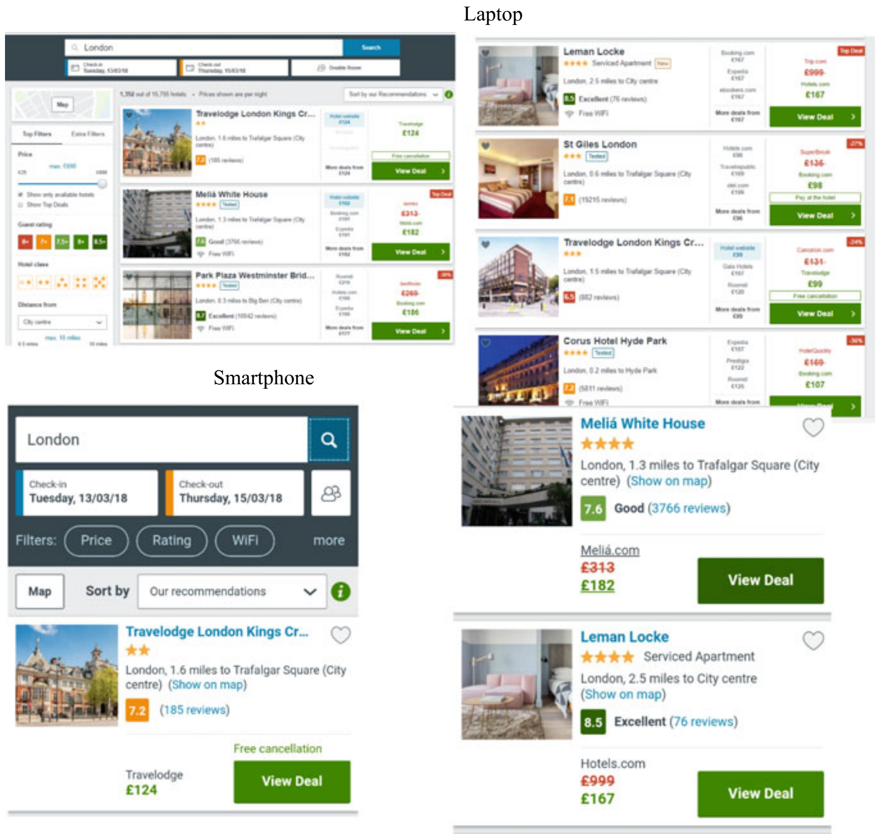


Fig. 3.8 Hotels displayed on www.trivago.com in smartphones or on laptops at the same time (information searched on 25 January 2018)

information gathered during this stage is very useful (Xiang et al. 2015) and can be exploited by big data to apply more effective marketing strategies.

3.5 Big Data Challenges in OTAs

This section summarizes the big data challenges faced by OTAs according to the model proposed by Sivarajah et al. (2017, p. 265) and considering the empirical research that forms the basis of this chapter.

Data *(Volume, Velocity, Variety, Variability, Veracity, Visualization and Value)*

The fact that OTAs are online businesses that generate a high volume of sales makes the investment in big data crucial and means that the volume management of data is one of main challenges. This research has proven that through the booking process and the customer experience, OTAs can collect a great deal of information in the format which best suits them thereby facilitating its exploitation. The information obtained from other sources, such as social media, may be available in other formats so that it is necessary to adapt such data before it can be used.

So that any information gathered be useful to analysts, it must be precise and consistent. Moreover, data must be updated and its meaning comprehensible. These challenges can be met by accurately defining the information requested from the customer on the webpage of the OTA, for example, by using drop-down lists and reducing the open fields, although the latter can be very useful to obtain qualitative information. In order to take advantage of the information from a comment or text a wide range of new software is appearing.

All this data must be managed with agility in order to make it available when it is needed. Doing so gives a company a competitive advantage by drawing valuable conclusions from big data. OTAs can decide how to display the most relevant information in the most attractive or user friendly formats, thereby improving the customer experience. This has led to the development of software that allows OTAs to be more responsive in their interpretation of the data by making it more readable.

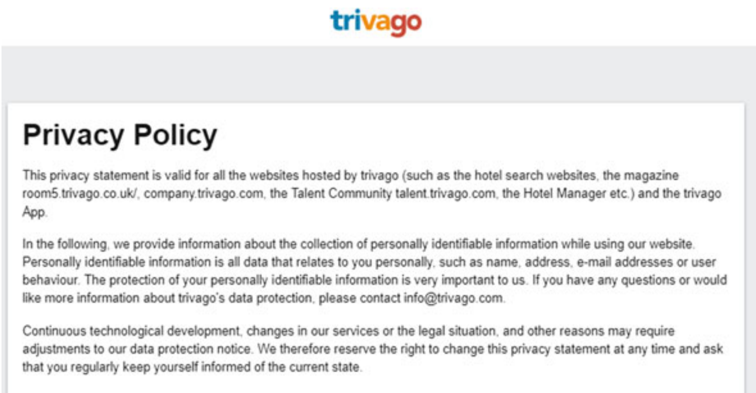
As a result, the investment in a variety of technologies is recommended, whatever the sector is.

Process *(Data Acquisition and Warehousing, Data mining and cleansing, Data Aggregation and integration, Analysis and Modelling, Data Interpretation)*

Data captured and acquired must be valuable and storable. To obtain the greatest amount of valuable data, OTAs have to properly define all the fields and filters that the user must or can fill on the webpage. Empirical research revealed the existence of more than 50 filters, some of which are personalized, which, in fact, can be useful for big data purposes. Regarding the data acquired it is very important that OTAs select which information is needed, to acquire it from reliable sources and in manageable formats.

Once all the data is gathered and ready to be extracted, a cleansing process is necessary to filter out the data which can be imprecise, unreliable, inconsistent or useless previous to its exploitation. Due to the high volume of data, the information extracted must be aggregated or integrated in order to make it more meaningful and useful. From that point onwards, data is analyzed and modelled and consequently data is interpreted for which it is very important to select employees with analytical skills in order to maximize its value to the OTA. This leads to a better knowledge of the Customer Journey Map which can result in new marketing strategies at each stage. In fact, some of the measures taken by OTAs include the adaptation of their webpages in order to display more convincing messages, to launch promotions more efficiently, to offer more useful information and to ease the booking process.

Management *(Privacy, Security, Data Governance, Data and Information sharing, Cost/Operational expenditures and Data Ownership)*



trivago

Privacy Policy

This privacy statement is valid for all the websites hosted by trivago (such as the hotel search websites, the magazine room5.trivago.co.uk/, company.trivago.com, the Talent Community talent.trivago.com, the Hotel Manager etc.) and the trivago App.

In the following, we provide information about the collection of personally identifiable information while using our website. Personally identifiable information is all data that relates to you personally, such as name, address, e-mail addresses or user behaviour. The protection of your personally identifiable information is very important to us. If you have any questions or would like more information about trivago's data protection, please contact info@trivago.com.

Continuous technological development, changes in our services or the legal situation, and other reasons may require adjustments to our data protection notice. We therefore reserve the right to change this privacy statement at any time and ask that you regularly keep yourself informed of the current state.

Fig. 3.9 Trivago privacy policy

As the volume of information gathered by OTAs is enormous and it can come from a diversity of sources, it is essential to protect customers' privacy (empirical research shows that all the webpages have a Privacy Policy as shown in Fig. 3.9) and to reduce their vulnerability to attacks by hackers aiming to steal their data.

To face up to all these challenges it is important that OTAs define a Data Governance policy. The tasks involved in the data governance include protecting users' rights and the OTAs ownership of the data, ensuring the security of the information and its quality, defining exchange of information policies, and analyzing the operational costs in order to be more effective and efficient.

3.6 Discussion

OTAs is a market that continues to grow and new competitors are continually appearing. From an historical point of view, Tripadvisor began as an advisory website and has become a meta-search engine. Most airlines offer accommodation, some of them establish partnerships with OTAs or meta-search engines, while others develop their own technologies (as is the case with Ryanair who recently created its own OTA) looking for more profitable options. Another potential competitor which is expected to have a significant impact on the market is 'Book on Google' that permits consumers to book accommodation without leaving Google.

OTAs can trace all user online activity from their first contact with the users to the customer satisfaction survey experience, that is, the whole customer experience. As a result, OTAs can take advantage of big data's possibilities from the first moment to develop new and more customized and disruptive marketing strategies that can benefit OTAs, establishments, customers, and tourist destinations. Furthermore, by using big data, OTAs could reduce the occurrence of what Anderson (2011) refers to as the 'billboard effect': users that search information in an OTA but finally book directly with the hotel.

Given that the number of bookings made via smartphone is on the rise, it is suggested that OTAs not only invest in the exploitation of big data, but also ensure that they present their services so that the information best suits each screen size, thus informing and persuading customers more effectively (Espinet et al. 2017; Espinet and Espinet 2017).

The use of big data is carried out and can be used internally in different ways.

- Identifying user patterns, habits, expectations and desires (Banerjee and Chua 2016).
- Defining more efficient pricing policies. Revenue or yield management strategies are developed using specific platforms for which data collection is of utmost importance. When using price discrimination policies, the use of some information can create risk, so it is vital to know the law and to act in accordance with it (Steppe 2017).
- Creating alerts and indicators that can be useful within various departments of businesses that use big data throughout the entire company. Some OTAs, for example Booking, Expedia, Logitravel, Trivago or TripAdvisor, have developed their own rating systems. Not only are they very useful internally, but also for users and for academic research (Stringam and Gerdes 2010; Liu et al. 2017; Mariani and Borghi 2018).
- Creating customer journey maps (CJM), which can be general but also segmented, using different criteria, such as the type of electronic device used by the consumer (Table 3.5).

An appropriate use of big data combined with services which have been customized to maximize the possibilities offered by each type of electronic device creates another huge range of opportunities for OTAs to broaden their businesses bases, and increase sales and profitability. Some examples are presented below.

- OTAs could try to use GPS and highly accurate tracking technologies to construct timeliness that could provide more complete information about what tourists do and where they visit during a stay in a particular destination so that OTAs can offer customers a more personalized range of options. This can be communicated through smartphones which are differentiated from the other electronic devices. These technologies are being used in different cities (Raun et al. 2016).
- OTAs could resell data to other businesses. Data collection, processing and management can be applied not only to their own business, but also to other commercial ventures.
- OTAs could develop commercial tools to help their suppliers, such as hotels. This is the case of Booking, TripAdvisor (called 'TripAdvisor Instant Booking') or Trivago (called 'Trivago Express Booking'). For the same purpose, Expedia created 'Expedia Media Solution' and Last Minute 'The Travel People'. In sum, the aim of these tools (which provide information regarding customer knowledge, spending habits, patterns of demand, price comparisons and so on) is to offer

Table 3.5 OTAs: sources of information useful for big data through the customer experience

Stage	Big data appliances	Differences depending on electronic device
Access OTA webpage	Cookies policy	NO
	Language Selection	NO
Registration (optional)	Fields requested	NO
Filters	The filters allow the application of more accurate marketing strategies	YES
Hotels displayed in the search results	– Hotels displayed – Order in which the hotels are displayed – Pricing policy – Characteristics and attributes – General information about availability	YES
Booking process	– Customer preferences – Complementary services booked	NO
Customer experience	Customer satisfaction	YES

The main difference in electronic devices is the quantity of information displayed

support to accommodation providers who may use the information to improve their marketing strategies.

- Some OTAs have opened new businesses or partnerships related to travel activities such as selling baggage, clothes, excursions and so forth.
- The data obtained could help OTAs develop consultancy services that could be used in different types of electronic devices.
- OTAs could develop partnerships with destinations or with suppliers, such as hotels, to build new strategies that benefit service providers as well as customers (e.g., Expedia establishes strategic alliances with hotel chains, such as Bahía Principe). In fact, OTA technologies have greater potential and reach than those employed by accommodation suppliers. Therefore, hotels should take advantage of their presence in the market and avoid direct competition with OTAs in areas where they have established their dominance. At this point, their relationship is characterized by codependence. OTAs identify new trends and requirements sooner; using this information, hotels can anticipate and meet customer demands sooner.
- OTAs could cooperate with other types of companies. For example, Destinia cooperates with Digitalmeteo to predict the effects of weather on bookings.

3.7 Conclusions

This chapter sheds light on the present and future of big data in OTAs with regard to the particular electronic device where the website or app is displayed. Use of big data is strategic for OTAs as it could allow these companies to gain a competitive advantage and reduce their costs. An adequate and efficient use of big data in OTAs can benefit not only these companies but also their customers, service providers and destinations in an ethical and responsible environment.

The results from the empirical research carried out specifically for this chapter reveal that OTAs use big data extensively throughout the entire customer experience. The main differences created by using different electronic devices is the quantity of information displayed due to the reduced size of the screens. Smartphones can provide OTAs another important difference through the use of GPS and highly accurate tracking technologies that enable these companies to obtain accurate information about what their customers do during a stay in a destination, so that these companies can offer a more customized service.

Big data can be very useful for the development of more efficient and personalized marketing strategies. Big data gives clues as to how information should be displayed in order to improve the customer experience, thereby persuading customers to spend more. In fact, new empirical research highlights the importance of design, the ease of use in the success of webpages (Bai et al. 2008; Smith et al. 2013). Information should be presented in a simple and intelligible way and the booking process should be simple which is why traceability is a key to the success of the website in this respect.

In spite of the extensive use of big data, untapped potential remains which could be exploited to derive further competitive advantage. This potential grows as the use of electronic devices, particularly mobile ones, grows. The main challenges facing OTAs are to be able to transform the data they collect into sustainable and useful information and to protect individuals' privacy in accordance with the law (i.e. the regulation of the market undertaken by the European Union as explained in Mantelero 2017). OTAs should prioritize their decisions regarding big data as an inefficient use of this asset could negatively affect their competitivity and cause resources to be wasted. In order to face up all their challenges and more efficiently dispatch all the tasks involving big data it is suggested that OTAs recruit experienced employees with business, IT and a knowledge of analytics, a resource which currently remains scarce.

To summarize, OTAs should consider big data as a mindset which affects the whole company and its organizational structure, and not only as information and its associated technology. Executives should align their big data aspirations—within the provided framework—with a realistic view of their own capabilities in order to get on the right track and stay ahead of the curve on innovation, competition, and productivity. It would be worthwhile to follow up this analysis in the future and it could be beneficial for OTAs to establish partnerships with academics so as to innovate further and undertake more advanced research.

Acknowledgements I would like to thank Laura Rafel, Àlex Espinet and Gethyn Rees for their help in this chapter.

References

Anderson CK (2011) Search, OTAs, and online booking: an expanded analysis of the billboard effect. Cornell Hospitality Report 11(8):4–10

Ayscue EP, Boley BB, Mertzlufft CE (2016) Mobile technology & resident attitude research. Tour Manage 52:559–562

Bai B, Law R, Wen I (2008) The impact of website quality on customer satisfaction and purchase intentions: evidence from Chinese online visitors. Int J Hospitality Manage 27(3):391–402

Baldwin H (2015) When Big Data projects go wrong. Forbes. Retrieved the 20 Jan 2018 from http://www.forbes.com/sites/howardbaldwin/2015/01/22/when-big-data-projects-go-wrong/#4b11e3a26231

Banerjee S, Chua AY (2016) In search of patterns among travellers' hotel ratings in TripAdvisor. Tour Manage 53:125–131

Cheng M, Edwards D (2015) Social media in tourism: a visual analytic approach. Curr Issues Tour 18(11):1080–1087

Côrte-Real N, Oliveira T, Ruivo P (2017) Assessing business value of Big Data Analytics in European firms. J Bus Res 70:379–390

Ditrendia (2017) Informe Mobile en España y en el Mundo 2017. Retrieved 25 Jan 2017 from http://www.amic.media/media/files/file_352_1289.pdf

Erevelles S, Fukawa N, Swayne L (2016) Big Data consumer analytics and the transformation of marketing. J Bus Res 69:897–904

Espinet JM, Espinet A (2017) Calidad y Satisfacción de los clientes del sector turístico a través de los dispositivos electrónicos: presente y futuro. Proceedings of International Marketing Trends Conference 2017 Madrid, 26–28 January, ed. Jean-Claude Andreani and Umberto Collesei, Paris-Venice Marketing Trends Association

Espinet JM, Espinet A, Filimon N (2017) Marketing through smartphones: the role of customer satisfaction and prices. In: 22nd International conference on corporate and marketing communications. challenges of marketing communications in a globalized world. 4–5, May 2017. ISBN: 978-84-946082-09

Gandomi A, Haider M (2015) Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manage 35(2):137–144

Guo Y, Barnes SJ, Jia Q (2017) Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet al location. Tour Manage 59:467–483

Hardy A, Hyslop S, Booth K, Robards B, Aryal J, Gretzel U et al (2017) Tracking tourists' travel with smartphone-based GPS technology: a methodological discussion. Inf Technol Tour 17(3):255–274

Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Khan SU (2015) The rise of "big data" on cloud computing: review and open research issues. Inf Syst 47:98–115

Kambatla K, Kollias G, Kumar V, Grama A (2014) Trends in big data analytics. J Parallel Distrib Comput 74(7):2561–2573

Khan N et al (2014) Big Data: survey, technologies, opportunities, and challenges. Sci World J 2014:18

Kim HH, Law R (2015) Smartphones in tourism and hospitality marketing: a literature review. J Travel Tour Mark 32(6):692–711

Li J, Xu L, Tang L, Wang S, Li L (2018) Big data in tourism research: a literature review. Tour Manage 68:301–323

Liu Y, Teichert T, Rossi M, Li H, Hu F (2017) Big Data for big insights: investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. Tour Manage 59:554–563
Mantelero A (2017) Regulation Big Data. the guidelines of the Council of Europe in the context of the European data protection framework. Comput Law Secur Rev 584–602
Mariani M, Borghi M (2018) Effects of the Booking.com rating system: bringing hotel class into the picture. Tour Manage 66:47–52
Mariani M, Di Felice M, Mura M (2016) Facebook as a destination marketing tool: evidence from Italian regional Destination Management Organizations. Tour Manage 54:321–343
Marine-Roig E, Clavé SA (2015) Tourism analytics with massive user generated content: a case study of Barcelona. J Destination Mark Manage 4(3):162–172
Miah SJ, Vu HQ, Gammack J, McGrath M (2017) A big data analytics method for tourist behaviour analysis. Inf Manage 54(6):771–785
Murphy HC, Chen MM, Cossutta M (2016) An investigation of multiple devices and information sources used in the hotel booking process. Tour Manage 52:44–51
Nikolopoulos K, Petropoulos F (2017) Forecasting for Big Data: does suboptimality matter? Computers and Operations Research, online
Oracle (2013) Big Data for the enterprise. OracleWhite Paper 1–16. Retrieved 20 Jan 2018 from http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf
Plaza B (2011) Google Analytics for measuring website performance. Tour Manage 32(3):477–481
Raguseo E (2018) Big Data technologies: an empirical investigation on their adoption, benefits and risks for companies. Int J Inf Manage 28:187–195
Raun J, Ahas R, Tiru M (2016) Measuring tourism destinations using mobile tracking data. Tour Manage 57:202–212
Sivarajah U, Kamal M, Irani Z, Weerakkody V (2017) Critical analysis of Big Data challenges and analytical methods. J Bus Res 70:263–286
Smith R, Deitz G, Royne MB, Hansen JD, Grünhagen M, Witte C (2013) Cross-cultural examination of online shopping behavior: a comparison of Norway, Germany, and the United States. J Bus Res 66(3):328–335
Steppe R (2017) Online price discrimination and personal data: a General Data Protection Regulation perspective. Comput Law Secur Rev 33:768–785
Stieglitz S, Mirbabaie M, Ross B, Neuberger C (2018) Social media analytics—challenges in topic discovery, data collection, and data preparation. Int J Inf Manage 39:156–168
Stringam BB, Gerdes J Jr (2010) An analysis of word-of-mouse ratings and guest comments of online hotel distribution sites. J Hospitality Mark Manage 19(7):773–796
Ur Rehman MH, Chang V, Batool A, Wah TY (2016) Big Data reduction framework for value creation in sustainable enterprises. Int J Inf Manage 36(6):917–928
Verma R, Stock D, McCarthy L (2012) Customer preferences for online, social media, and mobile innovations in the hospitality industry. Cornell Hospitality Q 53(3):183–186
Xiang Z, Schwartz Z, Gerdes JH Jr, Uysal M (2015) What can Big Data and text analytics tell us about hotel guest experience and satisfaction? Int J Hospitality Manage 44:120–130
Xiang Z, Du Q, Ma Y, Fan W (2017) A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. Tour Manage 58:51–65
Yaqoob I, Hashem I, Gani A, Mokhtar S, Ahmed E, Anuar N, Vasilakos AV (2016) Big Data: from beginning to future. Int J Inf Manage 36:1231–1247
Yovcheva Z, Buhalis D, Gatzidis C (2012) Smartphone augmented reality applications for tourism. e-Review Tour Res (eRTR) 10(2):2012

Chapter 4
Big Data for Measuring the Impact of Tourism Economic Development Programmes: A Process and Quality Criteria Framework for Using Big Data

Marianna Sigala, Andrew Beer, Laura Hodgson and Allan O'Connor

Abstract Big data revolutionalise the way organisations measure their performance and subsequently how they work. Technological advances allow organisations to access more data than they know how to handle and translate into value. However, although the literature has started investigating the use of big data for generating economic value, there has been a lack of research into the use of big data for delivering social value. To address these gaps, this chapter reviewed the related literature, in order to assist economic development agencies on integrating and using big data into their decision-making process and work related to the management of tourism economic development programs. To that end, the chapter develops and discusses a process framework for implementing big data initiatives and a decision framework for selecting and evaluating big data sources. The framework identifies four criteria for evaluating and selecting big data sources namely: need, value, time and utility. The implications of this framework for future research are discussed.

Keywords Big data · Decision-making · Performance measurement · Economic development programs · Process framework · Evaluation framework

4.1 Introduction

Big data is altering the way companies measure and monitor their performance. By changing the way organisations capture, analyse and use data, big data also transform the way organisations work and are managed. Thus, the widely mantra "*you are what you measure*", needs to nowadays be complemented with "…*you are also how you measure*". However, although a lot of studies have examined how big data is transforming for profit firms (Günther et al. 2017), limited research has paid attention so far on the use and impact of big data on the management and the work of governmental, non for profit organisations (Kim et al. 2014; Lavertu 2016) such as, economic development agencies.

M. Sigala (✉) · A. Beer · L. Hodgson · A. O'Connor
University of South Australia, Adelaide, Australia
e-mail: marianna.sigala@unisa.edu.au

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_4

However, by using big data for measuring, managing, selecting and evaluating the impact of economic development programs and activities, economic development agencies cannot only boost their economic value goals but also their contributions in generating social value. This is because by using big data for more effective decision-making and informed strategizing of economic development programs, economic development agencies do not only support individual economic actors, but they can also primarily alter the wider economies and societies in which they operate (e.g. by improving well-being, growing employment, supporting sustainability and liveability). In this vein, there is an urgent need to investigate the use of big data by public and not for profit organisations as well as to expand the literature by exploring how big data can be used for generating not only economic value (e.g. competitiveness, innovation, increased performance, better processes) but also social value for single users and the wider society as well.

In general, economic development agents assume, undertake and are responsible for various tourism development programs and activities, such as (EDA and Urban Enterprises 2018; IEDC 2016): the development and the management of the visitor economy; the attraction of investments and job creation; the enhancement of the well-being of local communities; and resource management and environmental sustainability. These functions are very similar to the roles and the responsibilities undertaken by Destination Management Organisations, DMO (Sigala and Marinidis 2012). Measuring the performance and the achievement of economic development agents and DMO has always been a difficult, if not impossible and challenging task (IECD 2014; Sigala 2014). Big data afford new measurement methods and metrics for addressing some of these challenges. Thus, it is important to explore how these organisations can better embed, collect and use big data for better informing their decision-making and strategic plans, as well as enhancing the effectiveness and implementation of their actions and practices.

Organisations nowadays have access to more data than they know how to analyse and what to do with it. Big data do not only refer to the vast amount of information that is nowadays captured and available through various technologies. Big data has also led to the creation of new technologies, methods, data capture applications, visualisation techniques and data aggregation capabilities. Thus, the traditional business analytic capabilities are not sufficient and appropriate to enable organisations to translate big data to value and make full use of it as an organisational asset. Instead, organisations need to develop new skills, capabilities, management styles and mindsets, as well as organisational structures and culture in order to turn big data into a useful resource and a competitive advantage. However, big data are not a panacea to any measurement problem. In addition, big data also entail many risks and limitations (e.g. privacy, ethical and accuracy/reliability issues) (Raguseo 2018) that organisations also need to consider when exploiting big data. Thus, the conversion of big data into value does not happen automatically; instead, organisations' readiness, openness and acceptability to appropriate big data are critically important factors that any big data exploitation project needs to also consider.

To address the abovementioned gaps, this chapter aims to investigate the use of big data by economic development organisations in order to better inform and manage the whole life cycle (i.e. selection, management, impact monitoring and evaluation) of tourism development plans and activities. In doing so, the chapter shows how big data can be translated into the generation of both economic and social value. To achieve that, the chapter first reviews the literature for identifying the information needs of these agents. Then, the characteristics, benefits and limitations of traditional and big data (sources and methodologies) are discussed in order to benchmark them against their capability to address the information needs of economic development agents. Finally, the chapter develops a framework that assists economic development agencies to identify and select what big data to use for informing their decision-making. The framework includes four criteria (need, value, time and utility), which consider the characteristics/features of big data. The framework was informed and developed based on the literature investigating the processes for exploiting big data as well as the criteria for selecting and evaluate open source data. The chapter concludes by discussing how economic development agents could implement this framework.

4.2 Economic Development Agencies: Functions, Information Needs and Measurement Trends

4.2.1 Functions

Economic development agencies perform numerous roles and functions that are very similar to the responsibilities assumed by DMO (Sigala 2014). Both organisations also share the same mission and goals, i.e. to develop and implement developmental strategies that contribute to the sustainable management and marketing of destinations/places. To achieve that, these organisations identify and support the implementation of economic development programmes with triple bottom line performance outcomes supporting the (economic, socio-cultural, psychological and physical) well-being of their communities and environmental sustainability of their places. For example, EDA and Urban Enterprises (2016, 2018) identified the following three overarching objectives of local economic development programmes pursued by local governmental agencies: supporting the existing business base; attracting new businesses and jobs; and, promoting liveability and sustainable communities. Tourism is a priority area and industry in which these organisations intervene for achieving their goals. Indeed, various core tourism economic development services are also found to fall under the responsibilities of Australian local governments (EDA and Urban Enterprises 2015), such as: tourism promotion and tourism industry development; industry support and development; destination development; policy development and advocacy; infrastructure creation; marketing and promotions; support for local business; information on demographics and trends; business efficiency network; economic and community asset network; new resident attraction; telecommunications; natural disaster recovery; skilled migration services; and careers forums for students.

4.2.2 Information Needs: Why to Measure and Why It Is not Measured

Similar to DMO and the importance to monitor their performance (Sigala 2014), measuring the impact of economic development activities undertaken by economic development agencies is critical. This is because the capacity to measure and demonstrate impacts of economic development programs both guides continuous improvement in the field and assists in securing resources to further growth at the local or regional scale. In addition, economic development agencies and DMO represent and 'have' to satisfy the needs of many stakeholders (e.g. elected members of council, taxpayers, (tourism) firms, governments at all levels, associations, trade unions), who have various (and sometimes conflicting) interests. Balancing the priorities and needs of these various stakeholders is not easy, while the provision of performance evidence is important for achieving the former. Thus, there is a pressing need to develop robust measures showing the efficient use of resources and taxpayers' money through the implementation and performance impact of economic development programs and activities. Lavertu (2016, p. 865) further explained the importance and urgency to measure the performance of governmental agencies by highlighting the benefits of peer learning and citizens' 'education': "… *public sector performance management—particularly the problem of goal displacement—the widespread dissemination of administrative data and performance information increasingly enables external political actors to peer into and evaluate the administration of public programs. The latter trend is consequential because external actors may have little sense of the validity of performance metrics and little understanding of the policy priorities they capture*".

Overall, these measures are needed in order to:

- justify the expenditure of scarce public sector resources by local governments, or indeed any tier of government;
- provide an evidence base on progress towards economic development goals;
- help local government agents to better allocate funds across economic development plans by monitoring and evaluating their impacts and ROI;
- assist in the management of economic development funds by assisting local governments to monitor their performance and impacts and take appropriate corrective actions;
- assist economic development practitioners select the most appropriate and effective strategies in order to achieve their economic development goals; and,
- build momentum in economic development efforts by creating knowledge of local success that can be shared with businesses, the community and other key stakeholder.

However, measuring the success of economic development activities has been challenging because of the difficulty—and high cost—associated with using conventional methods to assemble meaningful measures that reflect economic development outcomes. Similar to measuring DMO performance (Sigala 2014), measuring the impacts of economic development activities also suffer by the long-term horizon of various actions, the impact of various exogenous environmental factors influencing and/or interfering tourism activity and programs' results, and so, the inability to track future results and provide evidence of a causality between developmental actions and impacts.

Reality check studies (IEDC 2014) also show that despite the importance of performance measurement, more than 30% economic development agencies do not measure their performance regularly. Research findings also provided the following reasons why organisations do not track their performance. Reasons refer not only to the difficulties of measurement studies, but also to the deficiencies of data, methods and economic development agents to address measurement needs. Analytically, some developmental agencies did not measure, because there was disagreement over metrics and they lacked the resources—cash and staff time—to engage in effective performance measurement. Other impediments included uncertainty over which metrics were appropriate—or even usable—and concerns that key stakeholders may not fully understand the outcomes of tracking progress. Many agencies also acknowledged that many of the drivers of local growth are outside the sphere of influence of local economic developers, and therefore they were reluctant to measure something they could not control or substantially shape. The simple absence of appropriate data was a factor behind some agencies not track performance, while others felt that many development outcomes simply cannot be measured through quantitative indicators.

4.2.3 Information Needs and Measurement Trends: What and How to Measure

Recent studies (IECD 2016) reveal that the data most commonly sought and used by economic developers in the US include: labour regulation; employment by industry; taxes and incentives at the site; business revenues; building regulation; employers; demographics; higher education; and natural disasters. This is in contrast to the review of economic development performance approaches that was conducted by the IECD (2014) and which concluded that most non-profit organisations should measure four types of metrics: inputs; activities; outputs; and outcomes. It was also mentioned that the success of economic development programmes is equally dependent on both the measurement and the communication of the results of such activities. Thus, performance measurement should also include the customer satisfaction, which can be reflected in the measurement on how the target audience/stakeholders' views the relevance and helpfulness of an economic development agency (IECD 2014). The IECD's (2014) report also emphasised the need to adopt a 'balanced scorecard' method when assessing performance of economic development programmes. To that end, the measurement of 'soft' outcomes related to the environmental costs and benefits as well as of the Social Return On Investment (SROI) are critically important.

The IEDC (2014, p. 79) also identified a number of new approaches to performance measurement for economic development organisations, including:

- An internal assessment approach aiming to monitor and measure the internal processes that help the organisations conduct its business
- A relationship management approach focussing on creating and sustaining long term relationships with metrics tailored to capture how each stakeholder perceives the relationship
- A partnership approach measured by the degree to which organisations and stakeholders share values, beliefs and behaviours; thus, this approach builds and expands the relationship approach by aiming not only to balance but mainly to align the various stakeholders' interests
- A community approach aiming to foster and measure the stakeholders' involvement in economic development initiatives. Partnerships should be translated to collaborative networks demonstrating grassroots engagement in the development and measurement of economic development programmes.

However, it is not only the data needs of agencies and users within economic development that are continuously and rapidly changing. Recent reports also recognise that digital advances and big data can revolutionalise the measurement and the management of economic development programmes. Open, mobile and big data have not only increased the availability and accessibility of a vast of information, but they also offer the possibility of automated data collection (IECD 2016). Big data can also provide a new tool and lens for better managing the whole lifecycle of economic development programmes (Beer et al. 2018); from identifying priority areas and socio-economic problems to solve, to monitoring the performance progress and/or measuring the final impacts of economic development initiatives.

However, despite these technological advances and affordances, there has been little innovation in, or take up of, new data sources for economic development in the US (IECD 2016); where there has been innovation it has been solely in accessing open data from other government agencies. A recent study (Beer et al. 2018) identified the following barriers inhibiting economic development agencies to use big data in their operations:

- Limited skills in the manipulation and interpretation of big data, including familiarity with and capacity to use data sets and technology platforms/data sources
- Tightly constrained staff time to work on data analysis and the development of measures
- Budget restrictions for performance measurement
- Concerns over the reliability, security and ethical use of big data
- Government and agency silos restricting access to data.

Hence, there is an urgent need to investigate how economic development agencies can best integrate and use big data into their decision-making and operations, as well as develop tools in assisting them to use and translate data into socio-economic value.

4.3 Traditional Versus Big Data in Measuring Economic Developmental Programmes

4.3.1 Traditional Data

The impact of economic development efforts—are often difficult to develop, implement as well as measure (Beer et al. 2003; Turok 1989). Measurement difficulties are frequently attributed to the deficiencies of contemporary approaches to economic development measures, which suffer from a number of shortcomings:

- Official data from governmental bodies is often only available irregularly (e.g. population census), or at time periods that do not match the needs of the economic development agents. Moreover, the scale of this data analysis is often challenging as key data may only be available at a very broad scale (e.g. national level). These data limitations are usually due to the limitations of the data collection process.
- Primary data collection (i.e. the gathering of data for the specific purpose of evaluating economic development activities) is one solution, but attribution is an unavoidable problem (Turok 1989). Businesses often benefit from economic development actions but either overlook who provided that assistance, or are unaware of the boost they received. Moreover, the surveying of businesses or consumers can be expensive and often results in output measures, rather than outcome measures.

IECD (2016) summarised the limitations of conventional measurement approaches and metrics as follows: absence of standardised data collection; the mismatch between the spatial resolution of the data and the needs of the user; data being dated by the time it becomes available; the limited time quality and specificity of the data; the lack of representativeness; and the poor sampling.

4.3.2 Big Data: Benefits, Risks/Limitations and Requirements

In contrast to traditional data, economic development requires rich, multi-dimensional (economic, socio-cultural and environmental data from many stakeholders/perspectives), real time and spatially specific data. Big data are spatially and time based, as well as complex, i.e. generated from various users and sources. The characteristics of big data (Lehrer et al. 2018; Gandomi and Haider 2015) can significantly address the information needs for economic development measurement. Analytically, technology advances (sensors, social media, web-based tools) generate data in high volumes (large-scale data), at high velocity (high-speed real time data), in wide variety (data variability in the form of e.g. soft and hard data, text-based data and numerical data), and with a high level of veracity (multiple interpretations and a lot of 'noise', e.g. big data quality and reliability).

Big data have also the capacity to capture economic and social activity not measured by conventional and official statistics and metrics. For example, Uber data reflect transportation patterns and needs, while Airbnb data captures activities in black economy (e.g. number of micro-hoteliers and revenue generation) and economic development in residentials (non commercial) areas. Big data have the potential to complement and fill in the current gaps and deficiencies of official data and measurement methods.

Big data support citizen/stakeholder engagement in economic development performance measurement by crowdsourcing data collection. For example, by analysing twitter users' posts related to the status of corals in the Great Barrier Reef, Becken et al. (2017) explained how human sensors can be used for collecting data for environmental protection and management. Earlier, Sigala (2012) discussed how the two features of social media (namely social networks and user generated content) can be used for managing crises. These examples of crisis informatics and human/collective sensors for building crisis alert and performance measurement systems demonstrate the affordance and ability of big data to foster button up citizen engagement in economic developmental programmes and performance measurement activities.

Recently, Günther et al. (2017) identified two additional features of big data that can help organisations realize value from big data. These features (namely portability and interconnectivity) relate to the organisational context of big data use rather than their technological features. Thus, Günther et al. (2017) approach stresses the need to valorize and consider the socio-technical features of big data for generating value.

Portability refers to the possibility to transfer and remotely access big data from one context of application to be used in other contexts. Big data can be transferred and remotely accessed across technological platforms and organizational boundaries, and this technological affordance enables numerous organizational practices. Big data are captured once, but can be used multiple times and in various contexts. Thus, big data are not usually captured for the reasons for which they can be potentially used. For instance, big data captured by a transportation company like Uber can be used not only for identifying mobility patterns and needs, but they also reveal the most popular locations for various socio-economic activities (e.g. the most popular restaurants or bars in a city that can be used for destination marketing purposes). Portability, thus, seems crucial in the context of big data where the focus is on leveraging large volumes of varied data from many sources, considerably within, across but also beyond organizational boundaries. At an organizational level, organisations need to adopt a decentralized approach and management culture that will allow big data portability across functional silos. At work-practice level, professionals, analysts and decision-makers should be empowered to remotely access, transfer and use big data amongst platforms, technology applications, and institutional settings, without having a prior plan for collecting and using such data. At a supra-organizational level, organisations should develop and support open systems, culture and rules/legislation for enabling big data sharing and synergies across organizations.

The interconnectivity of big data enables users to go beyond the pre-existing templates of tapping isolated data sources by correlating and combining them in new ways (e.g. synthesized data are 'greater than the sum of its individual parts. The interconnectivity of big data needs to be exploited at all levels, but this in turns requires new capabilities and mindsets. At work-practice level, analysts and decision-makers need to be creative and adopt out of the box thinking for exploring connections and useful meanings amongst big data. To achieve that professionals need to maximize their ability to valorize the interplay between human and algorithmic intelligence. At an organizational level, interconnectivity require organisations to address tenses and trade-offs between centralized and decentralized decision-making as well as be more ready to accept business model change and transformation, e.g. Uber has been transformed from a mobility provider to a big data provider. At the supra-organizational level, interconnectivity needs to be valorized by nurturing and sustaining new partnerships and collaborations across organizations (even at an ad hoc 'plug and play' approach) but within the regulatory context of data privacy and security.

Overall, it becomes evident that: (1) big data afford many technological but also organizational capabilities to transform performance measurement and organization management; (2) big data use also entails several risks and limitations (i.e. privacy, proprietary information security, intellectual property risks, liability and quality issues) that need to be addressed when valorizing big data; and (3) big data translation into value requires a new set of skills, capabilities, organizational structures, policies and cultures at work-practice level, organizational level and supra-organisational level representing relations and dynamics with the institutional, political, competitive and technological environment of the organization.

4.4 Transforming Big Data into Socio-economic Value: Processes and Criteria for Selecting Big Data

4.4.1 Big Data Process Supporting Decision Making

Data analytics primarily consists of four stages, namely, data collection, data extraction, data warehousing, and knowledge generation. In this vein, big data processes require skills and capabilities referring to two major dimensions: data management and data analytics. Data manage require users to be able to identify what data to acquire and how; how to extract and clean data from inconsistencies and not related data; how to integrate, aggregate and represent visualize data (analyses) for making easier and better meaning. Data analytics refer to users' abilities and creativity to perform and identify appropriate data modeling analyses techniques as well as their ability to interpret data and identify appropriate actions and interventions. Recently, Braganza et al. (2017) developed and showed the practical value of adopting a strategic and archetype business process for developing and generating value from big data initiatives. This process framework consists of three action phases as follows:

Phase 1—commencement of big data initiative;
- Identify a strategic policy issue and establish benefits
- Develop and set the questions to be answered
- Establish a big data project team

– Determine and define the variables to answer the questions
– For each variable, identify indicators that can be measured
– Identify the sources of data for each indicator.

Phase 2—implementation of big data initiative;
– Acquire the data from the providers
– Specify the data (period of time, format etc.)
– Harvest the data to be analysed
– Agree and deploy analytical and statistical methods
– Synthesize the findings
– Visualise the information.

Phase 3—benefits from big data initiative.

– Decide actions to take action based on the information
– Implement the actors
– Manage the tangible and intangible results.

The focus of actions in phase 1 is clarification of definitions, terms, outcomes and likely results. This phase sets criteria by which decisions are taken and choices made during later phases, for instance, in relation to different combinations of resources, providers and methods of visualizing results. The big data protocol is particularly important where organizations plan to outsource their big data initiatives to third party businesses. This is because, one major reason for which big data outsourcing initiatives fail is because organizations enter into contracts with vague ideas of outcomes and actions they want from a big data supplier.

In phase 2 (big data implementation) users need to record, identify and select by justifying the benefits/costs of data analytics methods that they can use. Thus, at the core of phase 2 are decisions and choices made while conducting big data analysis. The creation of a 'trail of evidence' that sets out the rationale for choosing one set of options over another is critical during this phase.

Actions in phase 3 refer to translating data analytics into valuable and appropriate actions. However, people involved with and responsible for gaining benefits from big data are unlikely to be the same stakeholders involved in phase 2. This phase requires management commitment to ensure change is implemented over time. Outcomes of actions as a result of big data analysis need to be measured and recorded, so that lessons learnt from this initiative can be shared with future initiatives to improve the latter.

4.4.2 Criteria for Selecting and Evaluating Big Data

According to the Braganza et al.'s (2017) process framework for implementing big data initiatives, any big data project should start by identifying measurement dimensions and metrics that address the strategic targets and goals that the user needs to achieve and measure.

Thus, human intelligence, intuition and interests' dynamics are still be needed to identify and make a selection from the measurement priorities and areas, measurement metrics and methods as well as data sources that 'are' required. EDA and Urban Enterprises (2016) identified the various issues that users need to consider when selecting and judging data sets that can be used for measuring economic activity and outcomes. These key considerations include the need to select measures that:

- respond to the objectives of local government in economic development;
- are available at minimal or no cost;
- are available annually;
- are not time consuming to process;
- consider the diverse approaches to local economic development; and,
- are comparable across various local governments.

Research into open data can further help big data users on how to select, evaluate and use big data. Recently, Stróżyna et al. (2018) developed a framework for identifying, assessing and selecting open data sources. The framework identifies the critical steps that users need to go through for valorizing big data. The steps start by determining quality criteria for identifying data metrics, followed by identifying the potential sources of these data, the assessment of the data sources and finally the data retrieval process. The framework also stressed that data quality criteria for evaluating and selecting open data should be compatible with: (1) theory and data characteristics (e.g. accessibility, relevance, accuracy & reliability, clarity, timeliness & punctuality, and coherence & comparability); and (2) with the users' requirements and needs. The latter is in line with Braganza et al.'s (2017) big data process framework highlighting the need to start big data initiatives by determining and defining the users' strategic needs and priorities, which in turn big data need to address. ISO9000:2015 defines data quality as the degree to which a set of characteristics of data fulfills user's requirements.

Research within the open data context also highlights the need to start a data analytic initiative with the users' requirements and needs as well as determine the quality attributes of the data sets that one can use. Rula and Zaveri (2014) developed a comprehensive methodological framework for open data quality assessment that consist of the following steps: (1) Requirements analysis; (2) Data quality checklist; (3) Statistics and low-level analysis; (4) Aggregated and higher level metrics; (5) Comparison; (6) Interpretation.

Batini et al. (2009) also developed a process framework for using big data that also emphasises that a quality use of data should start and is determined by the user's perceived and identified data needs. This process framework consists of the following steps: (1) data analysis (examination of data schemas, complete understanding of data and related architectural and management rules); (2) data quality requirements analysis (surveying the opinions of users and experts to identify quality issues and set quality targets); (3) identification of critical areas (selection of databases and data flows); (4) process modeling (a model of the processes producing or updating data); (5) measurements of quality (selection of quality attributes and definition of corresponding metrics).

Quality and relevance of big data are important criteria, as big data are crowd-sourced (i.e. a lot of noise) and frequently collected for different purposes for which they might and can be used—i.e. problems of relevance. In the information systems literature, a lot of various data quality attributes can be found. In reviewing the literature Eppler (2006) identified 70 quality attributes of information and then narrowed the list to the 16 most important ones. Heinrich and Klier (2015) stressed to consider the quality of data as a multidimensional construct embracing multiple dimensions, e.g. precision, completeness, timeliness, consistency. Definition, methods and criteria for evaluating data quality are also very contextual determined. For example, data quality and importance of data quality attributes are perceived different in health, security and business domains (Braganza et al. 2017).

The European Parliament (2009) identifies seven data quality criteria: relevance (the degree to which data meet the current and potential needs of the users); accuracy (the closeness of estimates to the unknown true values; timeliness (the period between the availability of the information and the event or phenomenon it describes); punctuality (the delay between the date of the release of the data and the target date); accessibility and clarity (the conditions and modalities by which users can obtain, use and interpret data); comparability (the measurement of the impact of differences in applied measurement tools and procedures where data are compared between geographical areas, sectoral domains, or over time); coherence (the adequacy of the data to be reliably combined in different ways and for various uses). Later the ESS (2014) added the following dimensions of data quality: cost and burden (cost associated with the production of statistical products and burden on respondents); confidentiality (which concerns unauthorized disclosure of data); statistical processing (operations and steps performed by a statistical system to derive new information).

Research evaluating the quality of open data also reveals important dimensions of data quality that one has to consider when evaluating and selecting big data sets and sources.

Neumaier et al. (2016) stressed the possibility of open data sources to be linked with other sources as a critical open data quality. The metadata of open data (i.e. the data interpreting the open data), can compromise the searchability, discoverability, and usability of resources. Linkability of open data are found to be responsible for the low adoption and failure of open data projects (Neumaier et al. 2016). Similarly, the linkability of big data can significantly influence the interconnectivity and portability of big data, i.e. the two major features of big data as identified by Günther et al. (2017). Thus, linakability of big data should be considered as a major quality criterion.

Following this criterion, Neumaier et al. (2016) then specified meta-data quality issues in Open Data portals that are required in order to enable open data sources to interlink with each other. In reviewing the literature of data quality assessment methodologies applied to linked data, Zaveri et al. (2016) identified 18 quality dimensions and 69 metrics of linked data that are clustered into four groups: availability and licensing (accessibility dimensions); consistency and completeness (intrinsic dimensions); relevancy, trustworthiness, and timeliness (contextual dimensions); interpretability (representational dimensions). These dimensions are equally important for evaluating the quality of big data in order to ensure that users of big data can valorize the portability and interconnectivity affordances of big data.

Dorofeyuk et al. (2004) identified the following criteria determining the quality of a data source: understandability (a subjective criterion), extent (an objective criterion), and availability (an objective criterion), whereas the efficiency of a given data source is the weighted sum of its quality scores. Evaluating the sources of big data critically determines the quality of big data, and so, these criteria also need to be considered when selecting and evaluating big data.

Stróżyna et al. (2018) identified the following criteria determining the quality of open data sources:

- Accessibility—the possibility to retrieve data from a source; it includes such aspects as the structure of a source, the technologies used, the form of data as well as source stability (changes of a structure, errors, unavailability of a service); it also takes into account terms of use, privacy policy, requirements for login or registration, access to data (fees, subscriptions), etc.
- Relevance—what kind of information is provided by a source and whether this information matches the users' or system's requirements
- Accuracy & Reliability—reliability of information provided from the point of view of the users' requirements; it also evaluates data scope and coverage (how much information is available) and data accuracy (missing information)
- Clarity—availability of an appropriate description or explanation of the data model and information about a source of information (data provider)
- Timeliness & Punctuality—data update (time interval between the occurrence of an event and availability of the data which describe it) and time delay in publishing updated information
- Coherence & Comparability—whether data provided in a source describe the same phenomenon or have the same unit of measure like data from other sources.

4.4.3 Proposed Framework for Selecting and Using Big Data Sources for Measuring Tourism Economic Development Programs

The literature has shown that there are numerous dimensions based on which big data sources need to be assessed, selected and used. There is no big data set and source that can fulfil every criterion, while time and funds availability can be major constraints in terms of what data sets and sources an economic developer can use within a given context. Thus, it is useful for professionals to have a simple but simultaneously a robust decision framework to guide them about the value and suitability of big data sources and sets. To that end, a decision framework consisting of four dimensions and important quality criteria (Fig. 4.1), as identified earlier by the literature, was developed for assisting economic development agencies in their big data processes.

The decision framework identifies dimensions that if followed in the recommended sequence, it follows the big data process that user should adopt for implementing big data initiatives (e.g. Braganza et al. 2017).

The application of this decision framework first starts with the need, i.e. the identification of the purpose and the objective that economic development agencies need to achieve. What development programs need to be assessed, what are the objectives of these programs and what impacts they supposed to deliver? The specification of the purpose to be measured by using big data will then guide the identification of appropriate data sources and metrics that should be used for evaluating the materialisation of the programs' objectives. In other words, the first dimension of the framework refers to the need identified by the literature that big data processes and quality of big data need to start and consider the users' requirements and needs.

Once big data source needs and metrics are specified, the user need to assess required big data sources against their cost but always in relation to the benefits/functionality that he/she can get. Thus, the second dimension of the framework refers to value which relates to the costs, functionality/benefits and easy of use (accessibility) of big data sources. The scale and unit criteria of the second dimension capture the location and level of analysis benefits of big data, while the third dimension includes criteria that assess big data sources against the time benefits of big data in relation to their: timeliness (real time and/or past data); activity span (ability to specify and cover time periods of data capture); and repeated data (possibility for longitudinal data for time series analysis).

The last dimension includes criteria relating to: the interconnectivity and portability benefits of big data (e.g. ability to cross-utilise data in various tasks, functions and

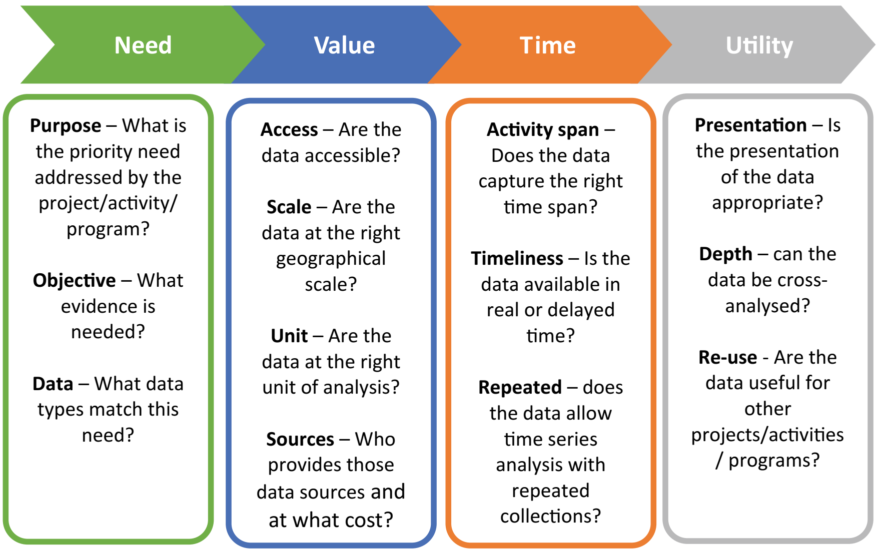


Fig. 4.1 Decision framework for assessing big data source suitability

possibility to replicate big data use in various contexts); and the big data interpretation process for understanding data and taking action (i.e. presentation and visualisation of data analyses for knowledge understanding, creation and extraction).

4.5 Conclusions and Implications for Future Research

Big data revolutionalise the way organisations measure their performance and subsequently how they work. Technological advances allow organisations to access more data that they know how to handle and translate into value. However, although the literature has started investigating the use of big data for generating economic value, there has been a lack of research into the use of big data for delivering social value. To address these gaps, this chapter reviewed the related literature, in order to assist economic development agencies on integrating and using big data into their decision making process and work related to the management of economic development programs. To that end, the chapters identified and discussed a process framework for implementing big data initiatives, which in turn also informed the development of a decision framework for selecting and evaluating big data sources.

The framework has not been tested and so, further research is required to tests its applicability and validity within a tourism economic development context. It is theorised that the selection of criteria for evaluating big data sources as well as their relative importance can be very much contextual dependent as well as stakeholder dependent. In other words, unit of analysis and location specific of data coming from Airbnb hosts' activity (who is the hosts and how much money they generate) may not be very important or required as a criterion when economic developers need to assess the impact of tourists' spending on a destination, but the former data will be important and necessary to break down at specific locations and profile of hosts when economic developers aim to assess the job creation impact of potential regulations on the Airbnb economy. Similarly, Airbnb data may be important for traditional players (e.g. hoteliers) but less important for micro-entrepreneurs and not relevant for stakeholders that are not affected by tourism. Thus, balancing stakeholders' interests and whose stakeholders' interests and needs the performance measurement of economic development programs is supposed to satisfy are also criteria that need to be considered when selecting and evaluating big data sources. In this vein, future research should be undertaken in order to investigate the impact of contextual and stakeholders' variables affecting the evaluation framework and its criteria. Future research should also be directed to examine appropriate strategies for managing conflicting and different stakeholders' interests as well as balancing the trade offs amongst the costs and the benefits of the various criteria.

Acknowledgements Support for this project was provided by Economic Development Australia with funding assistance from the Local Government Association of South Australia Research and Development Fund and in conjunction with the City of Adelaide, City of Salisbury, and the Eastern Region Alliance of Councils.

References

Batini C, Cappiello C, Francalanci C, Maurino A (2009) Methodologies for data quality assessment and improvement. ACM Comput Surveys (CSUR) 41(3):16

Becken S, Stantic B, Chen J, Alaei AR, Connolly R (2017) Monitoring the environment and human sentiment on the Great Barrier Reef: assessing the potential of collective sensing. J Environ Manage 203:87–97

Beer A, Haughton G, Maude A (2003) Developing locally. Polity Press, Bristol

Beer A, Hodgson L, O'Connor A, Sigala M (2018) Development and evaluation of economic development measures. Report prepared for Economic Development Australia (EDA)

Braganza A, Brooks L, Nepelski D, Ali M, Moro R (2017) Resource management in big data initiatives: processes and dynamic capabilities. J Bus Res 70:328–337

Dorofeyuk AA, Pokrovskaya IV, Chernyavkii AL (2004) Expert methods to analyze and perfect management systems. Autom Remote Control 65(10):1675–1688

Economic Development Australia, Victoria Committee & Urban Enterprise (2015) Local government industry performance monitoring and benchmarking survey

Economic Development Australia (EDA) & Urban Enterprises Victorian State Practitioners Network (2016) Annual performance measures of local economic development in Victoria. EDA, Melbourne, Victoria

Economic Development Australia, & Urban Enterprises—Victorian State Practitioners Network (2018) Best practice in economic development strategy: National survey results and discussion

Eppler MJ (2006) Managing information quality: increasing the value of information in knowledge-intensive products and processes. Springer Science & Business Media

European Parliament (2009) Regulation (EC) No 223/2009 of the European Parliament and the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom). Official J Eur Union 52

European Statistical System (2014) ESS handbook for quality reports. Eurostat

Gandomi A, Haider M (2015) Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manage 35(2):137–144

Günther WA, Mehrizi MHR, Huysman M, Feldberg F (2017) Debating big data: a literature review on realizing value from big data. J Strateg Inf Syst 26(3):191–209

Heinrich B, Klier M (2015) Metric-based data quality assessment—developing and evaluating a probability-based currency metric. Decis Support Syst 72:82–96

International Economic Development Council (2014) Making it count: metrics for high performing EDOs. IECD, Washinton

International Economic Development Council, IEDC (2016) A new standard: achieving data excellence in economic development. IEDC, Washington

Kim GH, Trimi S, Chung JH (2014) Big-data applications in the government sector. Commun ACM 57(3):78–85

Lavertu S (2016) We all need help: big data and the mismeasure of public administration. Public Adm Rev 76(6):864–872

Lehrer C, Wieneke A, Brocke L, Jung R, Seidel S (2018) How big data analytics enables service innovation: materiality, affordance, and the individualization of service. J Manage Inf Syst 35(2):424–460

Neumaier S, Umbrich J, Polleres A (2016) Automated quality assessment of metadata across open data portals. J Data Inf Qual (JDIQ) 8(1):2

Raguseo E (2018) Big data technologies: an empirical investigation on their adoption, benefits and risks for companies. Int J Inf Manage 38(1):187–195

Rula A, Zaveri A (2014) Methodology for assessment of linked data quality. In: 1st workshop on linked data quality, LDQ 2014, vol 1215. CEUR-WS

Sigala M, Marinidis D (2012) e-Democracy and web 2.0: a framework enabling DMOs to engage stakeholders in collaborative destination management. Tourism Anal 17(2):105–120

Sigala M (2012) Social media and crisis management in tourism: applications and implications for research. Inf Technol Tourism 13(4):269–283

Sigala M (2014) Evaluating the performance of destination marketing systems (DMS): stakeholder perspective. Mark Intell Plan 32(2):208–231

Stróżyna M, Eiden G, Abramowicz W, Filipiak D, Małyszko J, Węcel K (2018) A framework for the quality-based selection and retrieval of open data-a use case from the maritime domain. Electron Markets 28(2):219–233

Turok I (1989) Evaluation and understanding in local economic policy. Urban Stud 26(6):587–606

Zaveri A, Rula A, Maurino A, Pietrobon R, Lehmann J, Auer S (2016) Quality assessment for linked data: a survey. Semantic Web 7(1):63–93

Chapter 5
Research on Big Data, VGI, and the Tourism and Hospitality Sector: Concepts, Methods, and Geographies

Daniela Ferreira

Abstract The emergence of new data sources with high volume, velocity and variety has changed the market and consumption trends, including within the tourism industry. The tourism industry faces several challenges when using big data and volunteered geographic information (VGI) to develop new services. Managers need new capabilities to exploit this newly available information and to understand the possibilities that it has created. To give an overview of relevant academic knowledge, this chapter investigates academic research big data, volunteered geographic information, and the tourism and hospitality sector. It tracks the number of relevant publications and reviews the main concepts and research methods used in them. The results of this study give new insights on the main themes within scientific research on this topic, inform critical perspectives on the subject, and identify gaps in the current literature.

Keywords Volunteered geographic information (VGI), big data · Market research · Tourism

5.1 Introduction

In the big data age, large amounts of diverse information originates from three types of sources: direct, automated and volunteered information (Barnes 2013; Kitchin 2013, 2014; Kitchin and Dodge 2014). Volunteered information is especially valuable for consumption environments because it increases communication between consumers and producers. Many concepts have been proposed to address this issue: prosumption (Ritzer 2014), user-generated content; producer; civic hackers (Bruns 2008); crowdsourcing (Howe 2008; Haklay et al. 2008); and co-creative labour (Banks and Deuze 2009). Volunteered geographic information (i.e., information about the geographic location of individuals that they willingly share, such as by allowing a Google Maps app to access their smartphone GPS data) has a special value for tourism and

D. Ferreira (✉)
Centre for Geographical Studies, Universidade de Lisboa, Lisbon, Portugal
e-mail: danielaferreira@campus.ul.pt

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_5

hospitality because it is location-based and involves people that are not usually in the destination area. The many businesses that gather georeferenced information to enrich traditional maps with new content include Yelp, TripAdvisor, Zomato, Misk. Tourism businesses may have to adapt their business models to the resulting new possibilities for business innovation. Volunteered geographic information may not only transform market studies and strategies (González-Ramiro et al. 2016), but can also foster civic and critical reactions to tourism and its impacts on the environment and local communities (Ricker et al. 2012).

Although there have been previous reviews of smart tourism, (Gretzel et al. 2015), this chapter investigates scientific research up to 2017 through an analysis of academic publications about big data and volunteered geographic information in the tourism and hospitality sector, using a partly bibliometric approach (Koseoglu et al. 2016). The study counts the number of relevant publications found in the Web of Science and Google Scholar on this topic. It then reviews the main concepts and research methods used in the empirical research of the publications. The results allow a greater understanding of the routes that scientific research on this topic is taking, inform critical perspectives on the subject, and identify shortcomings in the current literature.

5.2 Digital Economy, Big Data and New Challenges

Technological advances generating big data have brought many challenges to the tourism industry (Sigala 2018a). New capabilities are required to process, analyse and manage the big data deluge (Fuchs et al. 2010; Townsend 2013).

Many digital platforms have emerged in the big data age to communicate and process digital information around a business model or theme. Firms are increasingly reliant on digital platforms to develop their services and products, and to create innovation. Platforms play an important intermediary role between firms and stakeholders, such as customers, online communities, suppliers, and advertisers. This new model of creating innovation has led to what has been called the new digital economy, platform economy or platform capitalism. This can be defined by the replacement of the industrial capitalism paradigm with the paradigm of economic digital networks based on the sharing economy (Langley and Leyshon 2017; Srnicek 2017; Pasquale 2016; McNeill 2016; Sigala 2018a). Platforms are integrated within ecosystems where there are online exchanges between different actors that allow co-creation of value and innovation (Yoo et al. 2016; Srnicek 2017; Sigala 2018b).

With the massive sharing and production of information by firms and users, the tourism industry faces new challenges. Platforms such as TripAdvisor, Booking, Trivago, Airbnb, with volunteered information, challenge firms with user reviews, classifications and other commentary (Yoo et al. 2016). Moreover, users are increasingly doing their hotel reservations, and choosing their favourite places, activities and other attractions, based on the shared experiences of other tourists. Therefore, platforms based on volunteered geographic information are central for the growth of tourism firms, creating new challenges for managers, regarding how they manage this huge amount of information (Townsend 2013).

The platform economy has been generating "reputation economies" and reputation functions as a form of capital, adding value to companies (Arvidsson and Peitersen 2013). Platforms play a key role in building reputational economies (Langley and Leyshon 2017). Users as consumers are increasingly important for reconfiguring products and services given the set of tools available online, including polls, ratings, and frequently asked questions. This gives direct and indirect feedback to firms (Callon et al. 2002; Piller and Walcher 2006). Thus user actions are increasingly important for economic growth (Grabher et al. 2008).

5.3 Methods

This chapter collects and reviews big data tourism research published in scientific journals or conferences proceedings. It encompasses editorials, original research (articles), review articles, short reports, book reviews, commentaries and conference papers. Books and book chapters were excluded because these are typically slower to appear. Clarivate Analytics Web of Science (WoS) and Google Scholar were used to identify relevant publications. WoS is the most used data base for bibliometric studies, along with Elsevier's Scopus. Google Scholar (GS) was used to widen the research coverage. WoS limitations include the predominance of English language sources, or the predominance of Natural Sciences, Engineering and Biomedical Research over Arts and Humanities and Social Science, which also apply to Scopus (Mongeon and Paul-Hus 2016), but these can at least partly be remedied through Google Scholar.

We searched the 'topic' field in WoS, which included the article's title, keywords, and abstract, but in Google Scholar, we only searched article titles because it does not permit a keyword or abstract search.

The first query used was 'Big Data AND Tourism'. This retrieved 261 results in WoS and 87 results in GS, with an overlap of 2 results. Many results were excluded due to (1) wrong publication types; (2) full-text not available online; (3) irrelevant content; and (4) no abstract in English. This left 96 articles, 70 from WoS and 28 from GS.

The second query was 'Big Data AND Hospitality'. The search retrieved 50 results in WoS and 17 results in GS, with an overlap of 2 results. After filtering out unwanted results, as above, there were 30 articles, 26 from WoS and 6 from GS.

After this wider analysis, we focused on understanding the importance of Volunteered Geographic Information (VGI) in the same sector. We conducted a third search with the key terms 'VGI AND Tourism' and 'VGI AND Hospitality'.

This query 'VGI AND Tourism' retrieved 8 results in WoS and 2 results in GS, with an overlap of 2 results. After excluding one result, there were 7 articles, 7 from WoS, 2 of which were also in GS.

The query 'VGI AND hospitality' retrieved no results from WoS and GS.

5.4 An Overview of the Publications

5.4.1 Type of Articles, Time-Series of Publications, Journals, and Affiliations

Although the first article about big data and tourism was published in 2011, the production of research within this field increased significantly only after 2015 (Fig. 5.1). This confirms that research about big data and tourism is recent. Most of the publications are original research articles (56) or conference papers (38). There is only one position paper, one research note and two editorials.

The big data and hospitality results are similar. There are 25 original research articles and 5 conference papers.

For VGI and tourism, the number of publications is scarce (7), and mostly published in 2016 and 2017, following the same trend as the previous queries. The type of publication is original research, except for one conference paper.

The conference papers are from 38 international conferences (Fig. 5.2). Most are from unique conferences, except for two IEEE (Institute of Electrical and Electronics Engineers) conferences with 3 publications each, and two conferences (IEEE and ASE—Academy of Science and Engineering) with two publications each. Despite this, some series of conferences have papers from distinct events, such as the IEEE International conference on Big Data or IFITT (International Federation for IT & Travel and Tourism) events. Conferences organised by the Institute of Electrical and Electronics Engineers seem to be the main forum for discussing issues related to the intersection of big data, tourism and hospitality, as 14 publications stem from these events. In fact, even beyond the IEEE events, most of the publications come from events organized by associations of engineering or technology. This in an important indication that a substantial section of big data tourism research has tended to be produced outside of the core tourism area, and probably by technology-focused researchers rather than by tourism researchers. The journal articles come from 48

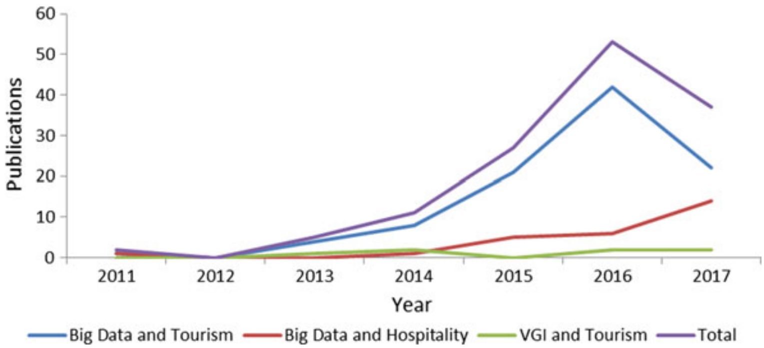


Fig. 5.1 Publications by year and query

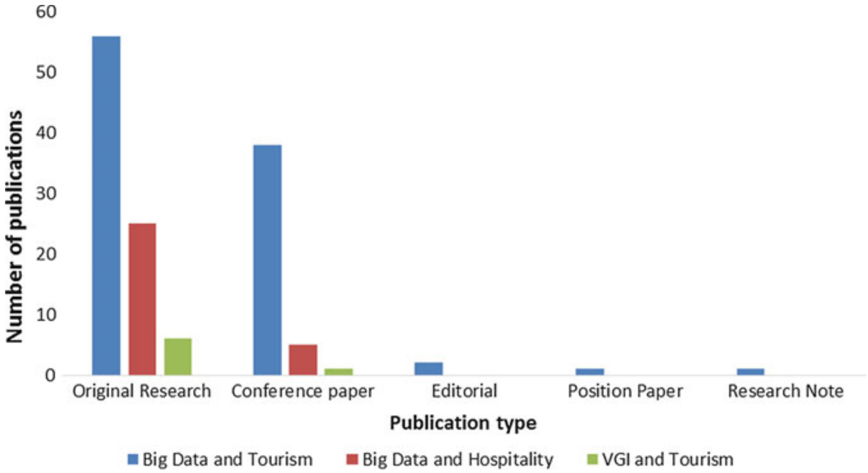


Fig. 5.2 Publications 2011–2017 by type

sources, and they paint a different picture of the sources of knowledge on the intersection between big data, tourism, and hospitality. The journal *Tourism Management* was the main forum for these issues, with 10 publications. It is followed by a series of journals in the field of tourism and hospitality management and marketing (*International Journal of Hospitality Management; Journal of Destination Marketing & Management; Journal of Travel & Tourism Marketing; International Journal of Contemporary Hospitality Management;* and *Journal of Travel Research*). The remaining journals vary in their main scientific field, but the field of tourism and hospitality research is in 19 journals of them. Other scientific fields present include information and data science, geography, management, engineering, science, economics and finance, and statistics. The dominance of engineering in conferences does not translate to a dominance of journals. Instead, journals in the field of tourism and hospitality form the main forum for issues about big data, tourism, and hospitality (see Table 5.1).

The scientific field of the first author of each publication was used to map the scientific fields involved in VGI/big data tourism research (Fig. 5.3). There are three main fields of research: tourism research, information and technology, and management and business. The first—tourism—appears as the main field with 22 publications, with additional publications under tourism and technology, tourism and administration, and tourism and economy. The second—information and technology—also gathers 22 publications, besides the related fields engineering, and business and technology. The cluster of management and business has a greater variety of affiliations. It includes management, business, business and management, and also the intersection of tourism and management, besides related fields such as marketing or economy and management. Other fields have less impact, but include mathematics and statistics, geography, social science, and spatial sciences.

Table 5.1 Journals with two or more publications 2011–2017

Journal	Articles
Tourism Management	10
International Journal of Hospitality Management	6
Journal of Destination Marketing & Management	5
Journal of Travel & Tourism Marketing	4
International Journal of Contemporary Hospitality Management	3
Journal of Travel Research	3
Annals of Tourism Research	2
Applied Geography	2
Asia Pacific Journal of Tourism Research	2
Cornell Hospitality Quarterly	2
Current Issues in Tourism	2
EPJ Data Science	2
Information Technology & Tourism	2
Information & Management	2
International Journal of Geo-information	2
Revista de la Facultad de Ingeniería U.C.V	2

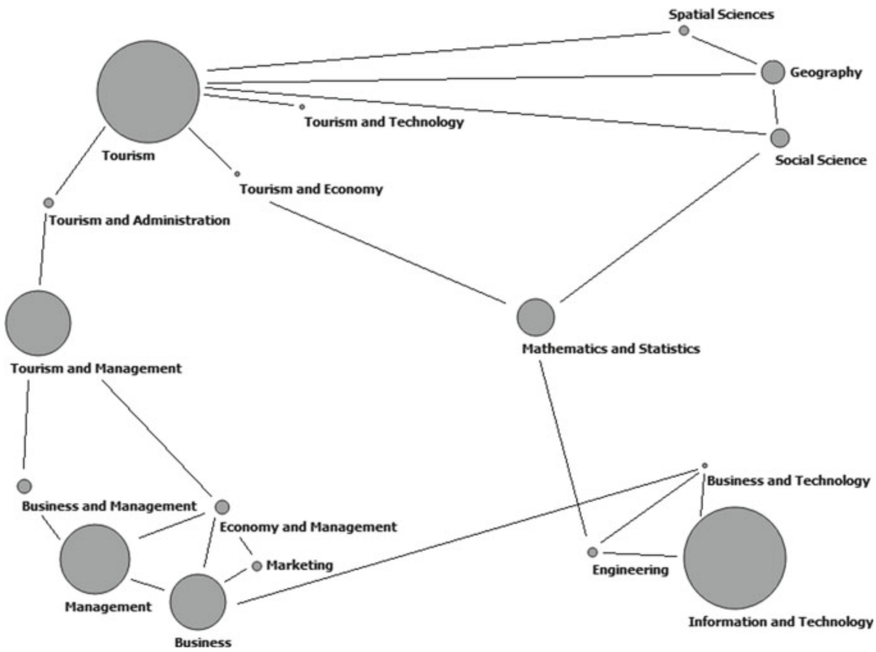


Fig. 5.3 Scientific fields of author affiliations for publications 2011–2017. Circle sizes correspond to the number of publications found. The location of each field and the lines connecting them represent their topic similarity

5.4.2 Geographies of Scientific Production

Publications about big data and tourism come from of academic institutions in China (39 out of 98), the United States (11), the United Kingdom (7), Italy (6), Spain (4), Australia (3), South Korea (3), Austria (3), Taiwan (3) and Germany (2). Seventeen countries have only 1 publication.[1] For big data and hospitality, publications are from United States (13), China (4), Spain (3) and Austria (2). Eight other countries have only one publication.[2] For VGI, United Kingdom and Italy have both 2 publications, while Germany, Spain and Canada have 1 publication each.

5.5 Themes, Main Concepts, and Research Methods

5.5.1 Big Data and Tourism

This section explores the main themes, concepts and methods that have emerged over the last few years within the field of big data and tourism in original research articles and conference papers. From the analysis of all publications, there are six big data and tourism research themes.

Creation of models or systems to support to decision-making and improvement of business (26 publications). This concerns how to use big data to improve strategies for the management and business of tourism. These works focus on predicting hotel demand or tourist arrivals, and on discovering popular tourist attractions to support the tourism industry. Important concepts include smart tourism and intelligent tourism. In relation to methods, many applications and quantitative models are discussed. The main issue that these models address is the necessity of taking advantage of large amounts of information that needs to be analysed.

Creation of models or applications for users to create and consume content (25 publications). This is characterised by the construction of models or applications to offer new possibilities for users to produce new content, generally also creating new big data analysis techniques. The underlying vision common to these studies is that it is necessary to develop new tools for data collecting, data management, computing and knowledge combination, that can take advantage of the new information that is being produced by users through social web tools. The main concepts are associated with big data analysis, cloud computing and system design. However, the concepts of smart tourism, smart destination, and social media, are significant in the context of creating new applications. These models and applications are assessed with quantitative approaches, including statistics and mathematical models.

[1] Romania; Sweden; Japan; Switzerland; Sri Lanka; France; Morocco; Brazil; Iran; India; Oman; Serbia; Estonia; Portugal; Indonesia; Ireland; Netherlands.

[2] Romania; United Kingdom; Switzerland; Australia; India; Serbia; Taiwan; Turkey.

The importance of user-generated content (20 publications). This includes research that focuses on the content that is produced by users, and especially tourists, through several platforms. These publications recognize the importance of reviews about hotels, restaurants and other tourist places. The articles investigate the preferences of tourists in each place, as well as customer attitudes/perceptions. This group discusses general platforms, such as Twitter, Flickr, TripAdvisor, travel forums and blogs, focusing on their potential for tourism. These host reviews about travel or other experiences, creating new challenges and new opportunities for tourism. The main concepts are co-production, user-generated content, social influence, business intelligence, and smart tourism. Sentiment analysis and text analytics as the predominant techniques in these publications. Other techniques include visual analytics, such as semantic pattern generation and network analysis. In the case of geotagged photos, spatial statistical techniques have been applied using Geographic Information Systems (GIS). In most publications, these techniques require large scale data collection (comments, reviews, photos) and therefore quantitative approaches are also necessary.

Literature review (15 reviews). The large number of reviews for this field reflects the need for a frameworks and overviews, including those that attempt to harness knowledge developed in other fields for tourism applications. For instance, the development of social web tools is important for tourism managers, generating questions about how it is possible manage this uncontrollable production of information and take advantage of it for their business. These publications identify new market trends and the best ways to respond to them. Commonly mentioned concepts include knowledge-intensive entrepreneurship, marketing knowledge, business intelligence, smart business ecosystems, business models, smart tourism, smart technology and open innovation. These concepts are closely associated with each other, and are likely a consequence of the big data age context that allows researchers to think about emerging concepts related to new automated modes of intelligence.

Tourism management and its new challenges (5 publications). These publications analyse customer-related data use and its applications. The central concepts are the big data age and e-commerce. The main techniques are qualitative: SWOT analyses and in-depth interviews with managers.

Bibliometric analysis (3 publications). These cover new data analytics, analysis of different perspectives of literature about big data in China, and analysis of business intelligence related to tourism.

5.5.2 Big Data and Hospitality

The ‘Big Data AND Hospitality’ search results were divided in four groups, most of which coincide with the groups identified in the query ‘Big Data AND Tourism’.

The importance of user-generated content (17 publications). Most of these studies go beyond the analysis of the content produced by users. While the publications in this section matching the query ‘Big data AND Tourism’ discussed the importance of several types of reviews for the tourism industry, the ‘Big Data AND Hospitality’ publications address deeper issues about the relationship between hotel guest experience and satisfaction, cultural and gender differences in the hospitality sector, and the factors that influence the frequency of travel and the influence on hotel performance. Some publications also focus on the structure of sharing economy. In sum, there is an increasing concern for understanding aspects of tourism and hospitality through user-generated online content. Most of the techniques used are very similar to the techniques used from the query ‘Big data AND Tourism’ but there are more interviews with hotel managers.

Tourism management and its new challenges (8 publications). These studies address the issue of improving management. As for the query ‘Big data AND Tourism’, the publications analyse the tendency of the use of data related to customers and its applications, as well as the identification of management challenges in respect to online content produced by users in some cases, and big data in general in others. Most of these studies use interviews with managers and other experts related with the sector in case studies.

Literature review (3 publications). These focus on statistical tests or the evolution of different perspectives and the new challenges and prospects for big data. In general, publications in this group approach the same topics referred in the group “Literature review” in the query ‘Big data AND Tourism’. They cover the need to understand big data, its evolution, challenges, and opportunities in a greater depth.

Others (2 articles). These cover the creation of models of big data analysis or models to support decision-making.

5.5.3 Volunteered Geographic Information and Tourism

The 7 VGI studies are similar to the group ‘The importance of user-generated content’ but are more focused on spatial analysis of the information collected in platforms that are based in cartography and geolocation. There are studies about the impact of VGI on users’ perceptions of online maps or studies about the evaluation of geolocation data (for example: photographs), using visual analytics and spatial analysis. The main difference is that the information has a predominantly geographic character.

5.6 Discussion and Conclusions

The results of the analysis of the big data tourism literature point to important research themes. Publications about big data and tourism often focus on the creation of models, systems and applications to improve big data analysis tools and to create improved platforms for users to post and consume online content. Research from China has played an important role in scientific production on this subject. There is an emerging necessity to understand big data challenges and opportunities (Li 2016). These studies focus on taking advantage of large sets of data to improve the tourism industry with new forms of analysis. Moreover, there is a concern about management strategies to respond to big data, given the changing management logic of most firms related with tourism or hospitality (Mariani et al. 2016; Vecchio et al. 2017). Online content is being produced in large amounts and the main concern of research is to understand how to exploit or control this free data.

For big data and hospitality, analysing user-generated content has been central in publications within this field. The priority is not so much to develop techniques of big data analysis to manage large amounts of information or to predict the number of potential tourists, but to analyse online content in greater depth. This gives opportunities to offer better spaces for content creation. The studies seek in-depth understandings about online content, including cultural and gender differences, and concepts such as the sharing economy.

Another research trend is conceptual development. The growing proliferation of informal and free information about tourist spaces and places has turned the user into an important digital influencer or opinion maker, which explains the increasing growth of works about this theme (McNeill 2016; Graham and Zook 2013; Elwood et al. 2012; Zwick et al. 2008).

Although there were few papers about VGI and tourism, VGI has a special value for tourism and hospitality due to the spatial contents. Given the strength of VGI to transform market studies and strategies (González-Ramiro et al. 2016), this subject is still underexplored by the literature.

Acknowledgements This study was supported by the Fundação para a Ciência e a Tecnologia under grant no. SFRH/BD/131253/2017.

References

Arvidsson A, Peitersen N (2013) The ethical economy: rebuilding value after the crisis. Columbia University Press, New York

Banks J, Deuze M (2009) Co-creative labour. Int J Cultural Stud 12(5):419–431

Barnes T (2013) Big data, little history. Dialogues Hum Geogr 3(3):297–302

Bruns A (2008) Blogs, Wikipedia, Second Life, and beyond from production to produsage. Peter Lang, London

Callon M, Meadel C, Rabeharisoa V (2002) The economy of qualities. Econ Soc 31:194–217

Elwood S, Goodchild M, Sui D (2012) Researching volunteered geographic information: spatial data, geographic research, and new social practice. Ann Assoc Am Geogr 102(3):571–590

Fuchs M, Scholochow C, Höpken W (2010) E-business adoption, use and value creation—an Austrian hotel study. J Inf Technol Tourism 11(4):267–284

González-Ramiro A, Gonçalves G, Sánchez-Ríos A, Joeng J (2016) Using a VGI and GIS-based multicriteria approach for assessing the potential of rural tourism in extremadura (Spain). Sustainability 8(11):1144

Graham M, Zook M (2013) Augmented realities and uneven geographies exploring the geolinguistic contours of the web. Environ Plan A 45(1):77–99

Gretzel U, Sigala M, Xiang Z, Koo C (2015) Smart tourism: foundations and developments. Electron Markets 25(3):179–188

Grabher G, Ibert O, Flohr S (2008) The neglected king: the customer in the new knowledge ecology of innovation. Econ Geogr 84(3):253–280
Haklay M, Singleton A, Parker C (2008) Web mapping 2.0: the neogeography of the GeoWeb. Geogr Compass 2(6):2011–2039
Howe J (2008) Crowdsourcing: why the power of the crowd is driving the future of business. McGraw-Hill, New York
Langley P, Leyshon A (2017) Platform capitalism: the intermediation and capitalisation of digital economic circulation. Finan Soc 3(1):11–31
Kitchin R (2013) Big data and human geography: opportunities, challenges and risks. Dialogues Hum Geogr 3(3):262–267
Kitchin R (2014) The data revolution. Big data, open data, data infrastructures & their consequences. Sage, London
Kitchin R, Dodge M (2014) Code/Space. Software and everyday life. MIT Press, Massachusetts
Koseoglu MA, Rahimi R, Okumus F, Liu J (2016) Bibliometric studies in tourism. Ann Tourism Res 61(1):180–198
Li Y (2016) SWOT analysis on the e-commerce system of Mt. Jiuhua tourism in the big data age. Account Finance Res 5(2):84–88
McNeill D (2016) Governing a city of unicorns: technology capital and the urban politics of San Francisco. Urban Geogr 37(4):494–513
Mariani M, Felice M, Mura M (2016) Facebook as a destination marketing tool: evidence from Italian regional destination management organizations. Tour Manag 54:321–343
Mongeon P, Paul-Hus A (2016) The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics 106(1):213–228
Pasquale F (2016) Two narratives of platform capitalism. Yale Law & Policy 35(1):309
Piller F, Walcher D (2006) Toolkits for idea competitions: a novel method to integrate users in new product development. R&D Manag 36:307–318
Ricker B, Johnson P, Sieber R (2012) Tourism and environmental change in Barbados: gathering citizen perspectives with volunteered geographic information (VGI). J Sustain Tourism 21(2):212–228
Ritzer G (2014) Automating prosumption: the decline of the prosumer and the rise of the prosuming machines. J Consum Culture 15(3):407–424
Sigala M (2018a) Market formation in the sharing economy: Findings and implications from the sub-economies of Airbnb. In: Barile S, Pellicano M, Polese F (eds) Social dynamics in a systems perspective. New Economic Windows, Springer, Cham
Sigala M (2018b) New technologies in tourism: From multi-disciplinary to anti-disciplinary advances and trajectories. Tourism Manage Perspect 25:151–155
Srnicek N (2017) Platform capitalism. Polity Press, Melden
Townsend A (2013) Smart cities: big data, civic hackers, and the quest for a New Utopia. W.W. Norton & Company Inc, New York
Vecchio P, Mele G, Ndou V, Secundo G (2017) Creating value from social big data: implications for smart tourism destinations. Inf Process Manage 54(5):847–860
Yoo KH, Sigala M, Gretzel U (2016) Exploring TripAdvisor. In: Egger R, Gula I, Walcher D (eds) Open tourism. Tourism on the verge. Springer, Berlin
Zwick D, Bonsu SK, Darmody A (2008) Putting consumers to work: co-creation and new marketing governmentality. J Consum Culture 8(1):163–196

Chapter 6
Sentiment Analysis for Tourism

Mike Thelwall

Abstract Sentiment analysis software is a key component of tourism big data research for its ability to detect positive and negative opinions in text. This supports large-scale analyses of the key affective dimension of reviews and social web posts about tourism venues and experiences. Sentiment analysis is fast and reasonably accurate, enabling patterns to be mined from large numbers of texts that would not be evident to experts reading those texts, such as differences between genders or venues in the aspects of destinations that are liked. This chapter reviews the main sentiment analysis approaches with a focus on practical descriptions of how the methods work and how they can be applied. The chapter also illustrates the value of sentiment analysis for tourism research.

Keywords Sentiment analysis · Tourism research · Social web posts · Online reviews · Tourism experiences

6.1 Introduction

Customer feedback and sentiment can help individual tourist attractions, hotels and restaurants to gain word of mouth recommendations and repeat visitors (e.g., Chen 2003) as well as to improve services (Schweidel and Moe 2014). Sentiment is at the heart of tourism because people expect to enjoy a holiday or visit. A critical part of the customer experience is therefore satisfaction: are they happy with the service that they received? When writing a review, the overall level of satisfaction may be flagged by an accompanying rating (e.g. 1–5 stars). Much feedback will not be explicitly rated, however, such as holidaymakers' tweets or Facebook posts (Philander and Zhong 2016). For big data analysis it is therefore essential to be able to detect the sentiment of this informal feedback automatically. This would reveal, for example, aspects of attractions that tend to appear in positive or negative comments.

M. Thelwall (✉)
Statistical Cybermetrics Research Group, University of Wolverhampton, Wolverhampton, UK
e-mail: m.thelwall@wlv.ac.uk

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_6

Sentiment analysis software has the task of detecting and classifying opinions and feelings expressed in text. Whilst there are many challenges to be solved before sentiment analysis programs attain expert human-level performance (Cambria et al. 2017), they give results that correlate positively with user ratings and so it is reasonable to use sentiment analysis scores when user ratings are absent (López-Barbosa et al. 2015). In addition, aspect-based sentiment analysis (see below) can give finer grained evidence than overall review ratings by identifying aspects of hotels or attractions that are singled out for praise or criticism. Sentiment analysis can also give additional insights into the important components of an offering. For example, an analysis of restaurant reviews found that context was important to diners in addition to those more widely recognised: food, service, price and ambience (Gan et al. 2017).

Sentiment analysis has become a standard component of the social media analysis toolkit of marketers and customer relation managers in large organisations (Hofer-Shall 2010). These may purchase relevant data (e.g., Facebook posts mentioning a hotel) and gain access to a suite of sentiment analysis and other tools to analyse it in a platform like Pulsar (pulsarplatform.com). Understanding how sentiment analysis works is therefore important for those working in tourism and the wider marketing sector.

This chapter reviews the main sentiment analysis approaches, describes how they may be exploited to analyse customer feedback and gives some examples from tourism research. It does not cover consumer sentiment indexes (Dragouni et al. 2016), but focuses on sentiment extracted from social media content, such as online reviews and comments (e.g., Tian et al. 2016). The chapter introduces the main current issues in sentiment analysis to the non-specialist and then reviews tourism-related sentiment analysis research. More technical details that are more relevant to computer scientists are avoided but can be found in previous reviews (Liu 2012; Pang and Lee 2008).

6.2 Core Sentiment Analysis Methods

6.2.1 Lexical Sentiment Analysis

The most transparent type of sentiment analysis is known as the lexical approach (Taboada et al. 2011). This uses a dictionary of sentiment-related terms, often together with estimates of their strength. The program SentiStrength (Thelwall et al. 2010; Thelwall et al. 2012), for example, includes 2846 positive or negative words or word stems (e.g., shabby, vex*). Each is accompanied by an estimate of its polarity and average strength in common usage. SentiStrength's positive terms are given a strength rating between 2 (mildly positive) and 5 (strongly positive). Negative terms are scored from −2 (mildly negative) to −5 (strongly negative). Thus, since *tremendous* scores +4 and *overheated* scores −3, SentiStrength would classify the text below as follows (i.e., containing both moderately negative sentiment −3 and strongly positive sentiment 4):

- Tremendous hotel but the rooms were overheated. → **−3**, **4**

As this shows, SentiStrength gives a separate positive and negative score. It can be tried online at http://sentistrength.wlv.ac.uk. Other programs give a single output (e.g., positive or negative).

Sentiment can be expressed in linguistically complex ways that a lexical sentiment analysis program might try to detect with additional rules. Most importantly, sentiment can be neutralised or flipped by negation and sentiment can be strengthened or weakened with booster terms like "very". For example, with SentiStrength

- I was not happy with the room but very satisfied with the view. → **−2**, **4**

Here, *happy* scores +3 but it is negated by "not", changing its value to −2. In contrast, "satisfied" normally scores +3 but is boosted by "very" to +4.

Sentiment can be expressed typographically with emoticons or with spellings that suggest enthusiasm. A lexical program might have a list of emoticons with sentiment scores and apply rules to detect enthusiastic spelling. SentiStrength would score "haaaapy" as +4 after detecting that (a) the underlying word is "happy" with score +3 and (b) the extra "aaa" increases the strength of this word.

- At the top of the Eiffel Tower:) → **−1**, **2**
- Haaaaapy to be in the Tivoli gardens! → **−1**, **4**

Given the relative simplicity of the above procedures it is important to check whether they work well enough in practice to be useful. This can be assessed by comparing computer and human scores for the same texts that have been implicitly or explicitly coded for sentiment by humans. An explicit coding might be derived from the sentiment judgements of expert coders whereas an implicit coding might be a rating associated with a review on the TripAdvisor website. Because computer accuracy varies by text type it is best to test the text type that will be analysed (e.g., hotel reviews) or multiple text types. SentiStrength has been tested on six social web sites, correlating with human judgements at a rate of between 0.30 and 0.65 for positive sentiment strength and between 0.50 and 0.60 for negative sentiment strength. Although these correlations are moderate, humans also tend to agree only moderately with each other. The agreement rate between SentiStrength and expert coders for negative sentiment strength is lower than the agreement rate between the expert coders but higher than for typical people doing the same task. The same is true for positive sentiment strength (except that it is less accurate on political discussions due to sarcasm).

As the above might suggest, sentiment analysis software can be expected to be less accurate than experts but more accurate than random people for the task of classifying a single isolated text. Sentiment analysis is fundamentally subjective, so humans often disagree. The primary advantage of software is speed: SentiStrength can classify 14,000 tweets per second but careful humans may require a minute to judge each tweet.

The above discussion described SentiStrength but all lexical sentiment analysis programs have their own lexicon and set of additional rules. Some, like Socal (Taboada et al. 2011), use advanced linguistic text processing to identify deeper structures in sentences, such as distinguishing adjectives from other words or disambiguating sentiment terms (Baccianella et al. 2010). This can improve their accuracy, especially on grammatically correct text, but reduces their speed.

6.2.2 Machine Learning Sentiment Analysis

Sentiment analysis can be achieved without a lexicon using machine learning (Liu 2012). For this, a set of human classified texts is needed. The algorithm learns how to identify positivity and negativity by examining many examples of both. To illustrate this, the program might notice that many negative reviews contain the term "dirty" and formulate the rule that future reviews are negative if they contain this term. Machine learning sentiment analysis typically focuses on the words, bigrams (consecutive word pairs) and trigrams (consecutive word triples) in each review, rather than complete sentences. Thus, a machine learning algorithm might use words and phrases like "love the room", "good service" and "rats" to help decide whether a hotel review is positive or negative overall. In practice, state of the art machine learning algorithms use complex pattern recognition methods, such as support vector machines (SVM) or deep learning, that are not understandable by humans.

- I <u>love the room</u>! We arrived early and took advantage of it. → **positive**
- We had <u>good service</u> from all the staff—prompt and polite. Thank you! → **positive**
- Could not believe we saw <u>rats</u> outside the restaurant kitchen. → **negative**

An advantage of machine learning sentiment analysis is that it does not require the human effort of creating a lexicon, in contrast to lexical sentiment analysis. In addition, its sentiment classifications can be more accurate than for lexical sentiment analysis if it has a large volume of training data (human classified texts) to learn from. An important disadvantage is that the machine learning sentiment analysis rules are opaque and so the causes of incorrect classifications cannot easily be deduced and corrected. Machine learning requires a substantial set of pre-classified texts (usually at least 1000) to achieve a reasonable performance. The results tend to be tailored to the type of texts used for training. Thus, a broad variety of pre-classified texts are needed to generate a generally useful algorithm.

For tourism-related reviews, all public automatic sentiment analysis tools seem to be less accurate than human judges for three different genres of tourism related text, with machine learning methods outperforming lexical software (Kirilenko et al. in press).

6.3 Universal Sentiment Analysis Tasks and Considerations

6.3.1 The Impact of Topic Domain on Algorithm Accuracy

A sentiment analysis algorithm designed for one language would not work well for another because it would not recognise the sentiment words. The same is true to a lesser extent for text types. A machine learning algorithm that learns how to identify sentiment from a set of book reviews might identify words like *interesting* and *readable* as positive. If it is then fed with the following hotel review it is likely to classify it incorrectly. In the illustrations below, the square brackets describe the type of sentiment algorithm applied to the data.

- Finding mouse poo under the bed was <u>interesting</u>. [books] → **positive**

In contrast, a machine learning algorithm training on hotel reviews is likely to have worked out that mentions of vermin or excrement indicate a bad experience and realise that the review is very negative.

- Finding <u>mouse poo</u> under the bed was <u>interesting</u>. [hotels] → **negative**

A machine learning sentiment analysis system trained on tourism related reviews has achieved an accuracy of 80% for the task of deciding whether they are positive or negative overall (Ye et al. 2009), presumably by learning a range of tourism-related sentiment expressions. Another system subsequently achieved an accuracy of 90% with a domain-specific lexicon (Gräbner et al. 2012). Thus, high levels of accuracy are possible for this task.

A general-purpose sentiment analysis program should be reasonably accurate on all types of text but perhaps not as accurate as one designed for the type of text being analysed. In contrast, a domain-specific program that is designed for one type of text is likely to be the most accurate for its domain. It is possible to develop methods to translate a program from one domain to another or to detect the domain of a text when processing it (Glorot et al. 2011). This has been shown to work well in the context of Tourism in the popular Chinese review site Ctrip.com (Li et al. 2015).

6.3.2 Language

Sentiment analysis algorithms tend to be language-specific because they learn from human coded texts from one language. Thus, for example, learning that “nice room” is positive does not help with identifying sentiment in Chinese hotel reviews.

- We had a <u>nice</u> room, which was a relief. [English] → **positive**
- We had a nice room, which was a relief. [Chinese] → **neutral**
- 我们有一个漂亮的房间, 这是一种解脱. [English]- → **neutral**

Although most sentiment analysis research has been conducted in English, there are many algorithms in other languages. For example, sentiment systems have been designed for Spanish TripAdvisor reviews (Salas-Zárate et al. 2017), Arabic TripAdvisor reviews (Cherif et al. 2016), and Russian Olympics-related tweets (Kirilenko and Stepchenkova 2017). A generic system has also been applied to Spanish TripAdvisor and Booking.com reviews (Fondevila-Gascón et al. 2016). More generally, there are multilingual toolkits, such as OpeNER, that have different language variants (García-Pablos et al. 2016). There are also methods to generate multi-lingual applications or to translate sentiment detection algorithm from one language to another (Balahur and Turchi 2014). These can be expected to be less accurate than language-specific versions.

6.3.3 Image Sentiment Analysis

Sentiment is often expressed explicitly in images (smiling for the camera) or implicitly (photograph of blue sky and golden sand) rather than in text. Images are increasingly used for informal communication between friends (Thelwall et al. 2016) due to the affordability and availability of smartphone image sharing apps as well as simple pathways to post images in social network profiles. It is much more difficult for computers to identify sentiment in images than in text because there is more data to process and it is more complex. Nevertheless, recent advances in image processing with convolutional neural networks (e.g., Oquab et al. 2014), which mimic the human brain to some extent, have led to breakthroughs with visual sentiment analysis. It is now possible to detect sentiment in some types of image with a high degree of reliability (You et al. 2015) but the process is much slower than for text processing.

Pictures have a central role in tourism (Chalfen 1979), from promotional images to the postcard home, souvenir snapshots and selfies at attractions posted to social media (Lyu 2016). Image-based sentiment analysis therefore has the potential to open an underexplored dimension of the leisure industry.

6.3.4 Universal Sentiment Analysis Tasks

Sentiment analysis programs have different output types. The most common is probably **trinary**: each text is classified as positive, negative or neutral overall. **Sentiment strength** or intensity is also sometimes estimated, as by SentiStrength (positive and negative separately) and Socal (combined positive-negative scale). Finally, **fine grained emotion detectors** attempt to distinguish between a set of emotions, such as happiness, sadness and anger (Neviarouskaya et al. 2009). Fine grained emotion detection seems to be not common due to the difficulty to obtain reasonably accurate results.

- Builders again at 5.30am! Aaaaargggggh!!!! [trinary] → **negative**
- Builders again at 5.30am! Aaaaargggggh!!!! [SentiStrength] → **−4**, **1**
- Builders again at 5.30am! Aaaaargggggh!!!! [SoCal] → **−3**: **very negative**
- Builders again at 5.30am! Aaaaargggggh!!!! [fine] → **anger**

Aspect-based sentiment analysis goes one step further by detecting both sentiment and the sentiment object (Jo and Oh 2011). Thus, it may detect in "cold shower but comfy bed" that the shower was bad [cold = negative] and the bed was good (comfy = positive).

- Cold shower but comfy bed → **shower**: **negative**; **bed**: **positive**

An aspect-based sentiment summarisation might list the aspects of a product or service that attract positive or negative feedback, generating useful market intelligence (Gamon et al. 2005; Reis et al. 2014). For tourism, this could take the form of identifying the positive and negative terms that are commonly associated with the key aspects of a hotel, such as its view (Marrese-Taylor et al. 2013a, b).

- [lots of reviews] → **view (positive)**: **sea**, **beach**; **view (negative)**: **distance**, **blocked**

In another case, a sentiment analysis of TripAdvisor reviews of California State Parks extracted opinions about aspects including the shops, camping, rangers, roads, shops and trails (Farhadloo et al. 2016).

One clever application of aspect-based analysis is to separate the text of a review into the different aspects of an attraction discussed in the review and then to apply different algorithms to the text associated with each aspect. This combines domain adaptation to improve the overall accuracy of the sentiment analysis results (Sharma et al. 2017).

6.3.5 Accuracy and Bias

As discussed above, automatic sentiment analysis is imperfect and not always correct. Ideally, these imperfections will average out and disappear when processing many texts. For example, a program might incorrectly classify the first text below as positive but over a large set of texts might work out that the hotel was upsetting residents by losing their post.

- The hotel lost my happy birthday card. → **positive**
- They lost my parcel and it is a disgrace. → **negative**
- I don't know how they lost my letter. → **negative**

It is impossible to eliminate all sources of bias but analysts should be aware that they exist and might influence the results.

Bias occurs in sentiment analysis when the errors are systematic in a way that would influence the conclusions drawn from the results. If a program persistently classified reviews of Happy Eater restaurant food as positive due to its name then an analyst might conclude that Happy Eater food was universally loved.

- We all had fish and chips at the Happy Eater. → **positive**
- I ate at the Happy Eater. → **positive**

Bias can also be subtler. For example, there are gender differences in the expression of sentiment (Teso et al. 2018) and sentiment analysis software is more accurate for female-authorised texts (Thelwall 2018a, b; Volkova et al. 2013). This is because there is a small but statistically significant tendency for females to express sentiment more clearly, as in the first of the two reviews below.

- Thanks so much to Sheila and Keith for keeping the bar open until 2am for us! All the girls had a wonderful time! → **positive**
- The hotel kept the bar open until 2am for the lads. What more can I say. → **neutral**

A big data analysis might give more weight to the opinions of females because it can detect them better. There can be important gender differences in the opinions of tourists (Yan et al. 2018), and so gender biases in detecting sentiment may affect the overall results. In contrast, an analysis of over 20,000 reviews of restaurants in two US cities found no gender differences in average sentiment (Micu et al. 2017), so gender bias may not always be relevant.

Bias can occur when one type of customer is more likely to post, such as younger users. Hotel customers might also be more likely to post if their positive or negative experience could be attributed to a named member of staff. For this reason, services might be commented on more than bed comfort even if both were equally important. In addition, positive experiences are more likely to lead to sharing on general social media sites whereas negative experiences are more likely to lead to posting a review to an integrated tourist site, such as TripAdvisor (Yan et al. 2018).

6.4 Special Considerations for Tourism-Based Sentiment Analysis

When applying or interpreting sentiment analysis to tourist reviews there are some important generic issues.

6.4.1 Limitations of Social Media Analysis for Tourism

As discussed above, an important limitation of any form of social media analysis is that the people that post to the social web form a self-selected subset of all customers and may be a highly biases subset. Customers that have had a very good or bad experience are more likely to post a review to a website and perhaps most likely if the experience was bad. Older people and young children may also be less likely to post an online review or share their experiences in social media because they are less likely to be web users. Conversely, busy parents may be frequently online but too busy to post reviews.

There may be more subtle biases in the demographics of users, such as in favour of some ethnicities or social classes. Some reviews may be fake—perhaps malicious reviews posted by competitors or paid positive reviews to boost a new or unpopular destination. Of course, self-selection bias also exists for most survey based research but this can be minimised by effective research designs that ensure high response rates. There is no equivalent for the social web and no good internet remedy for types of people that are not well represented online. Thus, researchers should be careful to identify sources of bias and their likely effect and interpret the results in the light of these.

Another limitation of online research is that it lacks evidence of direct connections with relevant actions, such as online bookings or purchases made because of the opinions expressed (Schuckert et al. 2015). Thus, whilst it seems reasonable to assume that negative reviews would not be good for an attraction, it is hard to quantify their impact in terms of reduced visits.

6.4.2 Tourism Domains

As discussed above, sentiment analysis programs can be general purpose or domain (i.e., topic) specific, with domain-specific applications likely to be the most accurate. In the context of tourism, a machine learning algorithm that had trained on mobile phone data might know that "expensive" is bad without learning whether more hotel-specific terms or phrases, such as "clean", "unfriendly" and "welcoming receptionist" expressed sentiment. Whilst a general-purpose sentiment analysis algorithm should give reasonably accurate results if enough customers use common generic sentiment terms (e.g., "excellent", "good value"), learning domain-specific terminology can improve performance. This even applies between different types of tourism offering. Whilst common sentiment terms for a restaurant would also apply to a hotel because hotels offer food, the reverse is true to a lesser extent. For other attractions, such as beaches, arcades and cultural activities, the key sentiment terms might be quite different.

- A sunny day but had expensive <u>mouldy</u> bread rolls and an <u>overheated</u> room. [hotels] → **negative**
- A sunny day but had expensive <u>mouldy</u> bread rolls and an overheated room. [restaurants] → **negative**
- A <u>sunny</u> day but had expensive mouldy bread rolls and an overheated room. [beaches] → **positive**

The key characteristics of reviews vary substantially between websites and so methods and findings from TripAdvisor or any other site should not be assumed to apply to all other review sites. For example, a comparison of TripAdvisor (439,000 reviews), Expedia and Yelp reviews in English of hotels found substantial differences in length (Expedia reviews were much shorter), themes (Yelp focused more on basic service) and sentiment (TripAdvisor and Expedia were dominated by positive reviews) (Xiang et al. 2017).

6.5 Applications of Sentiment Analysis to Tourism

TripAdvisor and other customer review sites seem to be taken seriously by many hotels and restaurants, who encourage customers to make positive comments and perhaps also respond individually to negative feedback. Positive reviews are recognised as important for future customers, many of whom book online with sites that feature customer reviews. Tourism managers may also analyse comments in non-review general social web sites, such as Twitter and Facebook.

6.5.1 Customer Relations Management Applications

Most national destination marketing organisations for the top ten destination countries mentioned social media analysis as a potential investigating tool in surveys conducted in 2010 (Hays et al. 2013) and today it seems likely to be an accepted tool. Providers of specialist services, such as development tourism (Jurowski 1998) may need to pay attention to online perceptions of their offering since negative sentiments could make it unviable. Of course, tourism organisations, attractions and hotels may run their own social media marketing campaigns (e.g., Hays et al. 2013) and may use sentiment analysis to help evaluate the success of individual initiatives. There isn't a definitive list of the most common applications of sentiment analysis to tourism but the following are likely candidates for inclusion.

- **Brand/business/service monitoring over time**: For example, a hotel chain might monitor its brand image in Twitter and Facebook daily and to identify long term trends in popularity. An early basic lexical sentiment analysis of an unspecified number of tweets from seven months in 2010 containing the words 'Phuket' or 'Bangkok' found some evidence of decreasing positivity (Claster et al. 2010).
- **Competitive intelligence**: A business can benchmark itself or investigate its competitors by analysing their social media presence in parallel with their own. For instance, a set of destinations could be rated and compared for sentiment (Valdivia et al. 2017a). This competitive intelligence might also identify market opportunities created by failures or threats posed by others' successful innovations.
- **Macro trends**: Tourism researchers can mine customer feedback and comments to identify broader trends than evident from individual hotels, restaurants, and attractions. For example, big data collections of tourist attraction reviews or other

feedback can help to identify which features are important for visitors (Alcoba et al. 2017). These may reveal patterns that are not evident from local information, such as the types of experience valued by different genders or nationalities.

- **Managing reputational risk**: Reputational risk management can be aided through the early online identification of surges in customer negativity (e.g., Das and Das 2016). The Hong Kong tourism boycott following the "occupy central" social movement was spread on social media (Sina Weibo), and could have been identified by routine monitoring (Luo and Zhai 2017).
- **Answer questions**: Managers may have specific problems or questions that can be investigated with social media data. An analysis of about 20,000 reviews about Paris in 14 categories (e.g., restaurants, shopping, off the beaten path) from virtualtourist.com. It used sentiment analysis to discover the main reasons why tourists were critical of the city transportation system (Kim et al. 2017).

6.5.2 Computer Systems to Support Tourists

Tourists can benefit from sentiment analysis through computing systems that provide recommendations. Software can do this by extracting the key aspects of potential destinations (e.g., hotels), extracting reviews about these destinations, processing the reviews with aspect-based sentiment analysis and then summarising the overall sentiment expressed about the customers potential choices. A web user wishing to decide between three hotels might be presented with a graphic showing their key aspects (e.g., breakfast quality, cleanliness, view) and then be shown the average rating of customers about those aspects (Schmunk et al. 2013). In theory, this would save customers the task of reading reviews if, for example, they were only concerned with one aspect of a hotel (e.g., children's entertainment, bar, beach) and so overall ratings that considered all aspects of a hotel would not be relevant. Whilst this would seem to be a useful service, there does not seem to be a successful commercial example so far. A promising but so far also apparently unsuccessful application is to overlay a map with icons representing tourism services, colour coded by sentiment (Cresci et al. 2014). Other applications have also attempted to exploit sentiment in tourist reviews to generate improved recommendations for them (e.g., Dong and Smyth 2016).

An unusual research application of sentiment analysis analysed the sentiment of tourism-related press coverage for countries and cities, generating an overall map of destinations by sentiment (Scharl et al. 2008). This study used a range of tools to gather the data and linguistic processing to identify relevant content and classify it for sentiment. The sentiment map may help customers to select destinations and plan itineraries.

An algorithm has been written to extract key phases from large sets of TripAdvisor hotel reviews to automatically summarise consumer feedback about them. Sentences were judged to be more relevant if they had been written by authoritative reviewers, were in reviews voted as helpful, were recent or contained a word or phrase indicating a thoughtful review (e.g., "all in all", "nevertheless"). Similar sentences were rejected so that the final set of key phrases represented differing perspectives (Hu et al. 2017).

Computer software can also recommend activities to users based on their interests. Using reviews of 1036 activities in the USA, one system attempted to cluster these activities by type and match them to user interests. Sentiment was harnessed to help ensure that poor attractions were not recommended (Mittal and Sinha 2017). This idea points to the benefits of using big data to help personalise the experience of individual users, although care must be taken to avoid confusing travellers with strange suggestions, however.

6.5.3 Research Insights from Review Sentiment

Some research studies have focused on the characteristics of reviews on the basis that they are important for businesses and that information about how they work may help businesses to interpret or react to them. A comparison of TripAdvisor review sentiment and ratings found them to be consistent for both budget and high-end hotels (Geetha et al. 2017). It also pointed to aspects of hotels that most affect reviewer sentiments. A statistical analysis of factors influencing the helpfulness of English language TripAdvisor reviews found that sentiment expressed in a review tended to make it more helpful to users, as judged by the number of helpful votes per month received by the review in the site (Hu et al. 2017). This study analysed all types of entity in three regions in the USA, with a total of over 700,000 reviews.

From a different perspective, an investigation of TripAdvisor reviews found that the overall rating of a review was frequently not a good guide to the sentiments expressed in the review, at least as judged by automatic sentiment analysis (Valdivia et al. 2017b).

An aspect-based sentiment analysis of over half a million TripAdvisor reviews of US Hilton hotels detected sentiment related to key aspects of a hotel stay (value, location, sleep, rooms, cleanliness, service, check-in, business service). The study also characterised reviewers by type and extracted review dates. From this volume of data, it was possible to compare average opinions about the different aspects of the hotels over time, between hotels, and between customer types, giving a rich dataset for exploration. The analysis also showed that the different types of customer gave different average ratings and therefore it could be misleading to compare average scores between hotels that attract different types of customer. Business users tended to be the least positive (Chang et al. in press).

Based on half a million TripAdvisor hotel reviews from New York City, visitors found negative sentiment to be more helpful in a review than positive sentiment (Lee et al. 2017). Potential customers may be more willing to discount positive reviews as potentially fake, but find negative reviews helpful to check whether there is an aspect of a hotel that would be unacceptable for them. Similar results were found in reviews of 10 Las Vegas hotels (Chang 2015).

Related to the above, an investigation of 676,000 tourism-related tweets about Milan from early 2011 found negative tweets to be more influential in the sense of being more likely to be retweeted (Barbagallo et al. 2012). This aligns with previous studies that have found negativity to be more powerful than positivity in provoking discussions (Chmiel et al. 2011) and so companies should be concerned to react to negative online comments.

Social media analysis can at least partially replace the traditional questionnaire strategy since it can exploit feedback already posted to social media. Whilst survey research has revealed that cleanliness is probably the single most important factor in hotel satisfaction and can analyse this issue in more detail (Lockyer 2003; Zemke et al. 2015), social media analysis has the potential to give even more fine-grained information through access to much larger samples of feedback.

An investigation into whether guided tours improve customer satisfaction at Spanish ports visited by Mediterranean cruises employed both questionnaires and TripAdvisor reviews, with sentiment extracted by the lexical algorithm Rapidminer 6.3. The sentiment analysis component showed that a guided tour helped to make each port visit a more explicitly positive experience, at least in terms of TripAdvisor comments. Alternatively, however, it is possible that more positive tourists are more likely to choose a guided tour (Sanz-Blas and Buzova 2016).

Many innovative applications of sentiment analysis have been proposed that are relevant to specific types of attraction. One application exploited geographic location information on Twitter to investigate the areas of Disneyland, California that attracted strong emotions, finding three areas with many positive tweets (Park et al. 2018).

6.6 Data Sources

There are many sources of tourism customer feedback. Structured websites like TripAdvisor are a good source of information because they contain explicit reviews and are categorised by attraction. This makes it easy to identify and interpret relevant data. General social media sites like Twitter, Weibo and Facebook can also be useful because they are probably used by a wider segment of the population. These sites are more difficult to exploit because heuristics are needed to identify relevant posts and the posts will not have explicit ratings. Businesses may also have their own data, such as responses to online feedback forms from their website and customer emails.

Twitter is an obvious general social web site for big data sentiment analysis because it is public and widely used to share information and updates. Although in the early days of Twitter, its data was distributed free to researchers via Spinn3r, this offering was withdrawn in 2010. Researchers can now either buy tweets from resellers or use the real-time free data collection option, the Twitter API. This allows tweets to be collected with a set of keyword queries but the data is restricted to the previous week. Large scale (millions) and/or long-term Twitter data can still be collected free for research projects that are planned well in advance. This is possible with free software like Mozdeh (mozdeh.wlv.ac.uk), COSMOS (socialdatalab.net/COSMOS) or Chorus (chorusanalytics.co.uk) to monitor collections of queries over a long period (typically months). Twitter's terms of service prohibit data sharing between researchers but analysis results can be published.

An investigation of sentiment related to Iizuka City in Japan is of interest for its method to identify relevant tweets. It used a manually curated list of 200 queries describing key attractions (e.g., "Kaho performing theatre" in Japanese) and an automatic query expansion method so that the resulting large set of Twitter queries could identify a substantial percentage of all relevant tourism-related tweets (Shimada et al. 2011).

Social media posts from Facebook and other widely used general purpose websites are probably more valuable than Twitter for businesses because they focus on sharing information and experiences between friends, giving a natural environment for posting about holidays. This data is rarely used by researchers because of the cost to access it, however.

6.7 Summary

This chapter introduced sentiment analysis, a mainstream commercial technique for analysing customer feedback and comments as part of a social media analysis strategy. Although sentiment analysis is imperfect, it is accurate enough to deliver useful information, such as identifying sentiment patterns and trends, when applied to big data. Standard sentiment analysis tasks include detecting the polarity of a text, the sentiment strength or fine-grained emotions expressed in it. Some software can also detect the aspect of a text that relates to sentiment expressed in it. A range of sentiment analysis software is available for applications but users should be aware of the two different types (lexical and machine learning) and the factors that can affect its accuracy, such as whether it is tailored for the tourism context.

There are many research-based and customer relations management applications of sentiment analysis. These exploit the sentiment expressed by consumers in their posts about an attraction, restaurant or hotel to mine insights that can inform decision makers and customer relations personnel. These vary from small scale issues like identifying the aspects of an individual hotel that trouble guests most to large scale identifying trends in customer reactions to a country or holiday destination region. Sentiment analysis can also be used to research theoretical issues related to customer satisfaction (e.g., whether guided tours improve the experience of visitors).

All users of sentiment analysis for big social media data should be aware of the self-selection bias limitations inherent in the medium. This is critical for researchers but also important for commercial users. A logical way to exploit social media data is to use it to identify issues or to get initial evidence to test hypotheses or explore data, with traditional methods, such as surveys, used for more robust checking of the most important points.

In terms of future research, a key attraction of sentiment analysis is that the free rich datasets currently extractable from websites like TripAdvisor allow big data approaches to be applied relatively cheaply to run larger scale studies than previously possible. This may give deeper and more general insights into tourism than previously possible. If recent attempts to extract sentiment from visual non-textual data, such as facial expressions (Soleymani et al. 2017), can be harnessed then future tourism research may also have access to new sources of implicit customer feedback that will support even more fine-grained analyses.

References

Alcoba J, Mostajo S, Paras R, Ebron RA (2017) Beyond quality of service: exploring what tourists really value. International conference on exploring services science. Springer, Berlin, Germany, pp 261–271

Baccianella S, Esuli A, Sebastiani F (2010) SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: LREC2010, pp 2200–2204

Balahur A, Turchi M (2014) Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. Comput Speech Lang 28(1):56–75

Barbagallo D, Bruni L, Francalanci C, Giacomazzi P (2012) An empirical study on the relationship between Twitter sentiment and influence in the tourism domain. In: Information and communication technologies in tourism 2012. Springer, Vienna, Austria, pp 506–516

Cambria E, Poria S, Gelbukh A, Thelwall M (2017) Sentiment analysis is a big suitcase. IEEE Intell Syst 32(6):74–80

Chalfen RM (1979) Photograph's role in tourism: some unexplored relationships. Ann Tourism Res 6(4):435–447

Chang WL (2015) Discovering the voice from travelers: a sentiment analysis for online reviews. Workshop on E-Business. Springer, Berlin, Germany, pp 15–26

Chang YC, Ku CH, Chen CH (in press) Social media analytics: extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. Int J Inf Manage

Chen JS (2003) Market segmentation by tourists' sentiments. Ann Tourism Res 30(1):178–193

Cherif W, Madani A, Kissi M (2016) A combination of low-level light stemming and support vector machines for the classification of Arabic opinions. 11th international conference on Intelligent Systems: Theories and Applications (SITA). IEEE Press, Los Alamitos, pp 1–5

Chmiel A, Sobkowicz P, Sienkiewicz J, Paltoglou G, Buckley K, Thelwall M, Hołyst JA (2011) Negative emotions boost user activity at BBC forum. Physica A: Stat Mech Appl 390(16):2936–2944

Claster WB, Cooper M, Sallis P (2010) Thailand—tourism and conflict: modeling sentiment from Twitter tweets using naïve Bayes and unsupervised artificial neural nets. In: Second international conference on Computational Intelligence, Modelling and Simulation (CIMSiM). IEEE Press, Los Alamitos, A, pp 89–94

Cresci S, D'Errico A, Gazzé D, Lo Duca A, Marchetti A, Tesconi M (2014) Tourpedia: a web application for sentiment visualization in tourism domain. In: Proceedings of the 9th Language Resources and Evaluation Conference (LREC 2014), pp 18–21

Das S, Das A (2016) Fusion with sentiment scores for market research. 19th international conference on Information Fusion (FUSION). IEEE Press, Los Alamitos, CA, pp 1003–1010

Dong R, Smyth B (2016) Personalized opinion-based recommendation. International conference on case-based reasoning. Springer, Berlin, Germany, pp 93–107

Dragouni M, Filis G, Gavriilidis K, Santamaria D (2016) Sentiment, mood and outbound tourism demand. Ann Tourism Res 60:80–96

Farhadloo M, Patterson RA, Rolland E (2016) Modeling customer satisfaction from unstructured data using a Bayesian approach. Decis Support Syst 90(1):1–11

Fondevila-Gascón JF, Mir-Bernal P, Puiggròs-Román E, Muñoz-González M, Berbel-Giménez G, Gutiérrez-Aragón Ó, Crespo JL (2016) Semantic fields to improve business: the hotels case. 5(2):47–59

Gamon M, Aue A, Corston-Oliver S, Ringger E (2005) Pulse: mining customer opinions from free text. Lect Notes Comput Sci 3646:121–132

Gan Q, Ferns BH, Yu Y, Jin L (2017) A text mining and multidimensional sentiment analysis of online restaurant reviews. J Qual Assur Hospitality & Tourism 18(4):465–492

García-Pablos A, Cuadros M, Linaza MT (2016) Automatic analysis of textual hotel reviews. Inf Technol Tourism 16(1):45–69

Geetha M, Singha P, Sinha S (2017) Relationship between customer sentiment and online customer ratings for hotels–an empirical analysis. Tour Manage 61(1):43–54

Glorot X, Bordes A, Bengio Y (2011) Domain adaptation for large-scale sentiment classification: a deep learning approach. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11). ICML Press, Bellevue, WA, pp 513–520

Gräbner D, Zanker M, Fliedl G, Fuchs M (2012) Classification of customer reviews based on sentiment analysis. Springer, Helsingborg, Sweden, pp 460–470

Hays S, Page SJ, Buhalis D (2013) Social media as a destination marketing tool: its use by national tourism organisations. Curr Issues Tourism 16(3):211–239

Hofer-Shall Z (2010) The forrester wave: listening platforms, Q3 2010. Forrester Res. http://www.demainlaveille.fr/wp-content/uploads/2010/07/forrester.pdf

Hu YH, Chen YL, Chou HL (2017a) Opinion mining from online hotel reviews–a text summarization approach. Inf Process Manage 53(2):436–449

Hu YH, Chen K, Lee PJ (2017b) The effect of user-controllable filters on the prediction of online hotel reviews. Inf Manage 54(6):728–744

Jo Y, Oh AH (2011) Aspect and sentiment unification model for online review analysis. In: Proceedings of the fourth ACM international conference on web search and data mining. ACM Press, New York, NY, pp 815–824

Jurowski C (1998) A study of community sentiments in relation to attitudes toward tourism development. Tourism Anal 3(1):17–24

Kim K, Park OJ, Yun S, Yun H (2017) What makes tourists feel negatively about tourism destinations? application of hybrid text mining methodology to smart destination management. Technol Forecast Soc Chang 123:362–369

Kirilenko AP, Stepchenkova SO (2017) Sochi 2014 olympics on Twitter: perspectives of hosts and guests. Tour Manag 63:54–65

Kirilenko AP, Stepchenkova SO, Kim H, Li X (in press) Automated sentiment analysis in tourism: comparison of approaches. J Travel Res

Lee M, Jeong M, Lee J (2017) Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website: a text mining approach. Int J Contemp Hospitality Manag 29(2):762–783

Li X, Li J, Wu Y (2015) A global optimization approach to multi-polarity sentiment analysis. PloS ONE 10(4):e0124672

Liu B (2012) Sentiment analysis and opinion mining. Morgan and Claypool, San Rafael, CA

Lockyer T (2003) Hotel cleanliness—how do guests view it? let us get specific. A New Zealand study. Int J Hospitality Manag 22(3):297–305

López Barbosa RR, Sánchez-Alonso S, Sicilia-Urban MA (2015) Evaluating hotels rating prediction based on sentiment analysis services. Aslib J Inf Manage 67(4):392–407

Luo Q, Zhai X (2017) "I will never go to Hong Kong again!" How the secondary crisis communication of "Occupy Central" on Weibo shifted to a tourism boycott. Tour Manag 62:159–172

Lyu SO (2016) Travel selfies on social media as objectified self-presentation. Tourism Manag 54(1):185–195

Marrese-Taylor E, Velásquez JD, Bravo-Marquez F, Matsuo Y (2013a) Identifying customer preferences about tourism products using an aspect-based opinion mining approach. Procedia Comput Sci 22:182–191

Marrese-Taylor E, Velásquez JD, Bravo-Marquez F (2013b) Opinion zoom: a modular tool to explore tourism opinions on the web. In: Proceedings of the 2013 IEEE/WIC/ACM international joint conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), vol 3. IEEE Computer Society, pp 261–264
Micu A, Micu AE, Geru M, Lixandroiu RC (2017) Analyzing user sentiment in social media: implications for online marketing strategy. Psychol Mark 34(12):1094–1100
Mittal R, Sinha V (2017) A personalized time-bound activity recommendation system. IEEE 7th annual Computing and Communication Workshop and Conference (CCWC). IEEE Press, Menlo Park, CA, pp 1–7
Neviarouskaya A, Prendinger H, Ishizuka M (2009) Compositionality principle in recognition of fine-grained emotions from text. In: Proceedings of the Third International ICWSM Conference (ICWSM2009). IEEE Press, Menlo Park, CA, pp 278–281
Oquab M, Bottou L, Laptev I, Sivic J (2014) Learning and transferring mid-level image representations using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Press, Los Alamitos, CA, pp 1717–1724
Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retrieval 2:1–135
Park SB, Kim HJ, Ok CM (2018) Linking emotion and place on Twitter at Disneyland. J Travel Tourism Mark 35(5):664–677
Philander K, Zhong Y (2016) Twitter sentiment analysis: capturing sentiment from integrated resort tweets. Int J Hospitality Manage 55(1):16–24
Reis G, Blair-Goldensohn S, McDonald RT (2014) U.S. Patent No. 8,799,773. U.S. Patent and Trademark Office, Washington, DC
Salas-Zárate MP, Paredes-Valverde MA, Rodríguez-García MÁ, Valencia-García R, Alor-Hernández G (2017) Sentiment analysis based on psychological and linguistic features for Spanish language. Current trends on knowledge-based systems. Springer, Berlin, Germany, pp 73–92
Sanz-Blas S, Buzova D (2016) Guided tour influence on cruise tourist experience in a port of call: an eWOM and questionnaire-based approach. Int J Tourism Res 18(6):558–566
Scharl A, Dickinger A, Weichselbraun A (2008) Analyzing news media coverage to acquire and structure tourism knowledge. Inf Technol Tourism 10(1):3–17
Schmunk S, Höpken W, Fuchs M, Lexhagen M (2013) Sentiment analysis: extracting decision-relevant knowledge from UGC. In: Information and communication technologies in tourism 2014. Springer, Cham, pp 253–265
Schuckert M, Liu X, Law R (2015) Hospitality and tourism online reviews: recent trends and future directions. J Travel Tourism Mark 32(5):608–621
Schweidel DA, Moe WW (2014) Listening in on social media: a joint model of sentiment and venue format choice. J Mark Res 51(4):387–402
Sharma S, Chakraverty S, Jauhari A (2017) Leveraging polarity switches and domain ontologies for sentiment analysis in text. In: International conference on computational intelligence, communications, and business analytics. Springer, Singapore, pp 84–92
Shimada K, Inoue S, Maeda H, Endo T (2011) Analyzing tourism information on Twitter for a local city. First ACIS International Symposium on Software and Network Engineering (SSNE2011). IEEE Press, Los Alamitos, CA, pp 61–66
Soleymani M, Garcia D, Jou B, Schuller B, Chang SF, Pantic M (2017) A survey of multimodal sentiment analysis. Image Vision Comput 65(1):3–14
Taboada M, Brooke J, Tofiloski M, Voll K, Stede M (2011) Lexicon-based methods for sentiment analysis. Comput Linguist 37(2):267–307
Teso E, Olmedilla M, Martínez-Torres MR, Toral SL (2018) Application of text mining techniques to the analysis of discourse in eWOM communications from a gender perspective. Technol Forecast Soc Chang 129(1):131–142
Thelwall M (2018a) Gender bias in sentiment analysis. Online Inf Rev 42(1):45–57
Thelwall M (2018b) Gender bias in machine learning for sentiment analysis. Online Inf Rev 42(3):343–354

Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A (2010) Sentiment strength detection in short informal text. J Am Soc Inf Sci Technol 61(12):2544–2558
Thelwall M, Buckley K, Paltoglou G (2012) Sentiment strength detection for the social Web. J Am Soc Inf Sci Technol 63(1):163–173
Thelwall M, Goriunova O, Vis F, Faulkner S, Burns A, Aulich J, Mas-Bleda A, Stuart E, D'Orazio F (2016) Chatting through pictures? A classification of images tweeted in one week in the UK and USA. J Assoc Inf Sci Technol 67(11):2575–2586
Tian X, He W, Tao R, Akula V (2016) Mining online hotel reviews: a case study from hotels in China. In: AMCIS 2016: surfing the IT innovation wave—22nd Americas conference on information systems, pp 1–8
Valdivia A, Luzón MV, Herrera F (2017a) Sentiment analysis in TripAdvisor. IEEE Intell Syst 32(4):72–77
Valdivia A, Luzón MV, Herrera F (2017b) Sentiment analysis on TripAdvisor: are there inconsistencies in user reviews? International conference on hybrid artificial intelligence systems. Springer, Berlin, Germany, pp 15–25
Volkova S, Wilson T, Yarowsky D (2013) Exploring demographic language variations to improve multilingual sentiment analysis in social media. In: EMNLP2013, pp 1815–1827
Xiang Z, Du Q, Ma Y, Fan W (2017) A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. Tour Manag 58(1):51–65
Yan Q, Zhou S, Wu S (2018) The influences of tourists' emotions on the selection of electronic word of mouth platforms. Tour Manag 66:348–363
Ye Q, Zhang Z, Law R (2009) Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Syst Appl 36(3):6527–6535
You Q, Luo J, Jin H, Yang J (2015) Robust image sentiment analysis using progressively trained and domain transferred deep networks. In: Proceedings of the twenty-ninth AAAI conference on artificial intelligence. AAAI Press, Menlo Park, pp 381–388
Zemke DMV, Neal J, Shoemaker S, Kirsch K (2015) Hotel cleanliness: will guests pay for enhanced disinfection? Int J Contemp Hospitality Manage 27(4):690–710

Chapter 7
Location-Based Social Network Data for Tourism Destinations

Konstantinos Vassakis, Emmanuel Petrakis, Ioannis Kopanakis, John Makridis and George Mastorakis

Abstract Social media networks are a resource for valuable knowledge about tourist destinations through the collection of data by Location-Based Social Networks (LBSN). A major problem is the lack of knowledge in respect to the visitors' views about a destination, as well as the fact that the visitors' behavior needs and preferences are not visible. Many enterprises and local authorities are still using traditional methods for acquiring knowledge to make strategic decisions, by collecting data from questionnaires. Nonetheless, this process, despite its benefits, is short-lived and the number of the participants is small compared to the number of visitors. This chapter discusses a methodology for the extraction, association, analysis, and visualization of data derived from LBSNs. This provides knowledge of visitor behaviors, impressions and preferences for tourist destinations. A case study of Crete in Greece is included, based upon visitors' posts and reviews, nationality, photos, place rankings, and engagement.

Keywords Big data · Tourism · Tourist destination · Social media · Location-based social networks · Data visualization

7.1 Introduction

The emergence of Big Data and the progress of data science have changed the way companies and organizations observe, analyze and influence the market. The global datasphere has been predicted to grow from 16.1 zettabytes (ZB) generated in 2016 to 163 ZB by 2025, unlocking unique user experiences and new opportunities. More than a quarter of this data will be in real-time and more than 95% of this will be produced by real-time Internet of Things (IoT) devices (Reinsel et al. 2017). Big data can be characterized by seven Versus: (1) *volume*, large datasets generated through

K. Vassakis (✉)
University of Crete, Rethymno, Greece
e-mail: kvassakis@gmail.com

E. Petrakis · I. Kopanakis · J. Makridis · G. Mastorakis
TEI of Crete, Heraklion, Greece

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_7

technologies, like IoT and user-generated content (UGC) from online social media platforms, (2) *variety*, the diversity of data sources and formats, (3) *variability*, data with meaning that can vary significantly in context, (4) *velocity*, high speed data generation through the growth of interconnected devices generating data in real-time, (5) *veracity*, data reliability, as data is worthless if it's not accurate, (6) *visualization*, visual representation of data and information in a way that is easily readable, accessible and understandable (Vassakis et al. 2017). Big Data is also recognized as creating (7) *value*, since it contributes to more efficient and effective operations like optimal price setting, optimizing supply chain, minimizing errors and improving customer satisfaction (Günther et al. 2017).

Tourism destination management is constantly transforming due to changing customer preferences/needs and technological developments. Big Data, structured and unstructured, should be harnessed to inform strategic decisions and enhance destination competitiveness, but the field is not yet well developed (Miah et al. 2017). In particular, understanding tourists' preferences and factors related to the location and services offered to maximize their satisfaction can support tourism destination management (Floris and Campagna 2014). In this chapter, data from Crete (Greece) from location-based social networks is analyzed, providing guidelines for local authorities and enterprises.

7.2 Social Media Analytics and Tourism

The social web hosts many interactions between customers with businesses. The spread of social media has turned the market online into a conversation (Chen et al. 2012). Interactions between companies and customers have substantially increased the data available, since it is possible to collect and evaluate a user's online activities, from their social media presence to entries in forums or customer reference sites to their online shopping behaviors. Companies and their market analysts can create a personalized profile for every customer and potential customer. The feasibility of getting these insights and how well these generated profiles can match with customers depends on the availability of data, technology, and analytics knowledge. The rise of mobile devices like smartphones, tablets, and laptops adds location-based information to the profiles of customers. This makes it possible to tailor marketing efforts to the preferences of people, with their physical position information bringing a variety of advantages to marketers and stakeholders (Vassakis et al. 2017).

Social media analytics is "*the analysis of structured and unstructured data collected from various social media channels*" (Vashisht and Gupta 2015). The most widely used social media platforms for analysis are social networks (e.g. Facebook and LinkedIn), microblogs (e.g. Twitter and Tumblr), media sharing (e.g. Instagram, Flickr and YouTube), social news (e.g. Digg and Reddit) and review sites (e.g. Foursquare and TripAdvisor). Analyses can be divided into two types: (a) content-based applications where the text and its language are the most significant factors for identifying users' emotions, preferences etc. and (b) structure-based applications where users' hobbies, interests, and relationships are clustered into communities (Chen et al. 2014).

Most tourism processes and transactions (from trip planning, bookings to tourist feedback) are digital. The bulk of tourists and travelers use the web and social media for travel planning and acquiring trustworthy information for their travel destination (Yoo et al. 2016). Therefore, an enormous amount of data around customers is generated at tourism destinations, identifying preferences and needs. Real-time analysis of social media data is a major driver for value creation in many industries (Vashisht and Gupta 2015; Vecchio et al. 2017). For example, Park et al. (2016) examined perceptions of Asian restaurants on Twitter. Using text mining, word frequency analysis and sentiment analysis on 86,015 tweets over four months, they found that the sentiment scores of Chinese restaurants were significantly lower than others. The most positive tweets were about food quality, while negative tweets suggested problems about the service quality or food culture.

There are several empirical studies examining user-generated content from social media in tourism destinations. Chang et al. (2017) used TripAdvisor to extract and visualize ratings and reviews for Hilton hotels. They exploited sentiment analysis and natural-language processing. They found the types of travelers that provided the lowest and highest ratings (business travelers and couples), the months with the lowest and highest rates (July and December) and the travelers' emotions according to the most frequently used negative or positive words.

Floris and Campagna (2014) developed a methodology that integrates data from TripAdvisor and Booking.com to extract meaningful knowledge for tourism planning and decision-making. They investigated tourist preferences, such as the most popular destinations, reasons that people chose those destinations and what they appreciate/ignore. They found that the success of a tourism destination depends on the quality of the tourism industry and the territorial setting of the destinations. Marchiori and Cantoni (2015) examined the impact of the prior experience of a destination and the change in users' perceptions following exposure to user-generated content. Using a dataset of 2505 American Internet users, they found that people who are more educated or have previously visited a destination are less likely to change their opinions after being exposed to online social media content. Marine-Roig and Clavé (2015) highlighted the significance of Big Data analytics for smart destinations, examining the online image of Barcelona with a dataset of 100,000 online reviews written in English by tourists. They claimed that their analysis provided significant guidelines for the stakeholders, leading to better strategy, marketing and branding for the tourism destination.

7.3 Location Intelligence and Tourism

The increasing number and availability of location-acquisition technologies including GPS, Wi-Fi and 4G allow users to publish media content (texts, photos and videos) along with their position as location-tagged media content, transforming social networks to Geosocial or Location-Based Social Networks (LBSNs) (Rathore et al. 2017).

Location Intelligence (LI) or Locational Data Analysis is "the use of locationally-referenced information as a key input in business decision-making" (Shekhar and Xiong 2007). LI expands Business Intelligence (BI), aiming to harness data for making strategic decisions, adding a spatial perspective to data analysis that creates critical context to the decision-making process with the incorporation of powerful data correlation and visualization methods (Milton 2011).

LBSN does not just add location as a new feature to an existing social structure so that people in the social network share information with a position tag, but also creates new social structures of individuals linked by interdependence derived from their physical world locations (Zheng 2011). In a LBSN, a user's location is represented as a place, such as a street, shop, park, beach, point of interest or building (Chorley et al. 2015) that is tagged in media content, like a post with text, picture or video providing information, to other users of the network informing them about the location of that post. Location-based social networks include Facebook, Foursquare, Google+, Instagram, Twitter, and Flickr.

Data-driven maps and location-based applications created using LI reveal spatial relationships and correlations with other types of business data that otherwise may not have been visible. Combining this data with other types of geographic data such as population, traffic and weather provide opportunities for analyzing various spatially-referenced phenomena and gain knowledge for businesses, organizations and public authorities. The analysis of data in both structured and unstructured forms, as well as the evolvement in data storage, data processing, and data mining technologies helps businesses and local authorities in making data-driven decisions and generating non-obvious knowledge in real-time (Ravi et al. 2018; Vecchio et al. 2017; Günther et al. 2017; Hashem et al. 2015). Businesses can use location data to understand the significant impact of "where" in their operations. LI as a part of the whole data analysis process allows businesses or public authorities to understand better external characteristics and how these affect their activities and to gain a comprehensive overview of a phenomenon by integrating location and time dimensions with internal data (Milton 2011).

Several prior studies have exploited location-based social networks. Lee et al. (2011) used a dataset of geo-tagged Twitter messages from Japan to monitor geographical areas, finding that crowd activity can reveal expected and unexpected events. Brandt et al. (2017), using 600,000 Twitter messages in San Francisco, examined the potential value of the spatial and semantic analysis of social media messages for smart tourism ecosystems. They found that social media analytics can reveal spatial patterns within the city related to presence, environmental and topical engagement, and these patterns contribute to value creation for smart urban tourism.

Using social media networks, tourists post an enormous number of photos. Photos with geotags that are published on photo-sharing social media networks like Instagram and Flickr provide opportunities for transforming information into knowledge. Content that is accompanied with photos offers rich information about tourist preferences and experiences. Zhou et al. (2015) proposed a method to find tourist hotspots through public Flickr images, demonstrating with the United States.

7.4 Location-Based Data for Tourism Destinations: The Case of Crete

Harvesting real-time spatial data from social media networks such as Twitter, Instagram and Flickr is challenging due to the volume of data. A data analytics system is therefore needed that processes offline data efficiently within a time limit and provides real-time data analysis for various social networks, including Twitter (Rehman et al. 2014; Carvalho et al. 2017; Komorowski et al. 2018; Liu et al. 2018), Flickr (Li et al. 2018; Donaire et al. 2014), Instagram (Schmidbauer et al. 2018; Mittal et al. 2017) and Foursquare (Liu et al. 2016; Wang et al. 2015). This case study examines questions related to tourists' preferences, such as which is the most popular destination (Floris and Campagna 2014) for the two largest cities of Crete (Heraklion and Chania).

For the data analysis process, the system using public user-generated content from the LBSNs consists of four main sub-processes: (1) data acquisition, (2) data cleansing and storage, (3) data querying and filtering and (4) data visualization (Fig. 7.1). The four selected LBSNs (Instagram, Flickr, Foursquare and Twitter) can be divided into two main categories, Media Sharing and Text Posts & Reviews, based on the types of data they generate. In the data acquisition process, the system communicates with the LBSNs repeatedly through their RESTful Web Services (RESTful APIs) to collect all the new generated data in an unstructured form, such as post text and language, image URLs, likes, shares, comments, locations, dates and times, usernames, photos, descriptions and followers. After acquisition, data cleansing and transformation is needed to select the required data and transform them in a structured form to be stored in the SQL database. The data from the main SQL database table (holding all the collected and transformed records) is also transferred to a data warehouse in the cloud. Querying a table in a data warehouse is fast since it is optimized for analytic access patterns and processes highly complex queries overall data (Bouadi et al. 2017). Afterward, data from each executed query is filtered to be visualized (Ferreira et al. 2013; Mahmud et al. 2016).

One of the major research challenges is the translation of enormous amounts of data into a comprehensible form for managers (Chen et al. 2012). Data visualizations are a good tool for this (Huang et al. 2017). The data was obtained from the LBSNs used in this study for two months (November to December 2017), as summarized in Fig. 7.2.

Data visualizations can reveal popularity and sentiment for specific days and points of interest. Positive sentiment dominates during the period examined, while the negative posts are in low levels (Fig. 7.3). The spikes in the graph also point to dates on which users were mainly positive or negative about the attractions or events offered. The most positive posts and reviews were about the tastefulness and quality of traditional food, the beauty of various points of interest, while negative posts were mainly about service quality issues (Fig. 7.4).

There were fewer textual posts from Twitter and Foursquare than photos from Instagram and Flickr for both cities. Twitter was more widely used than Foursquare

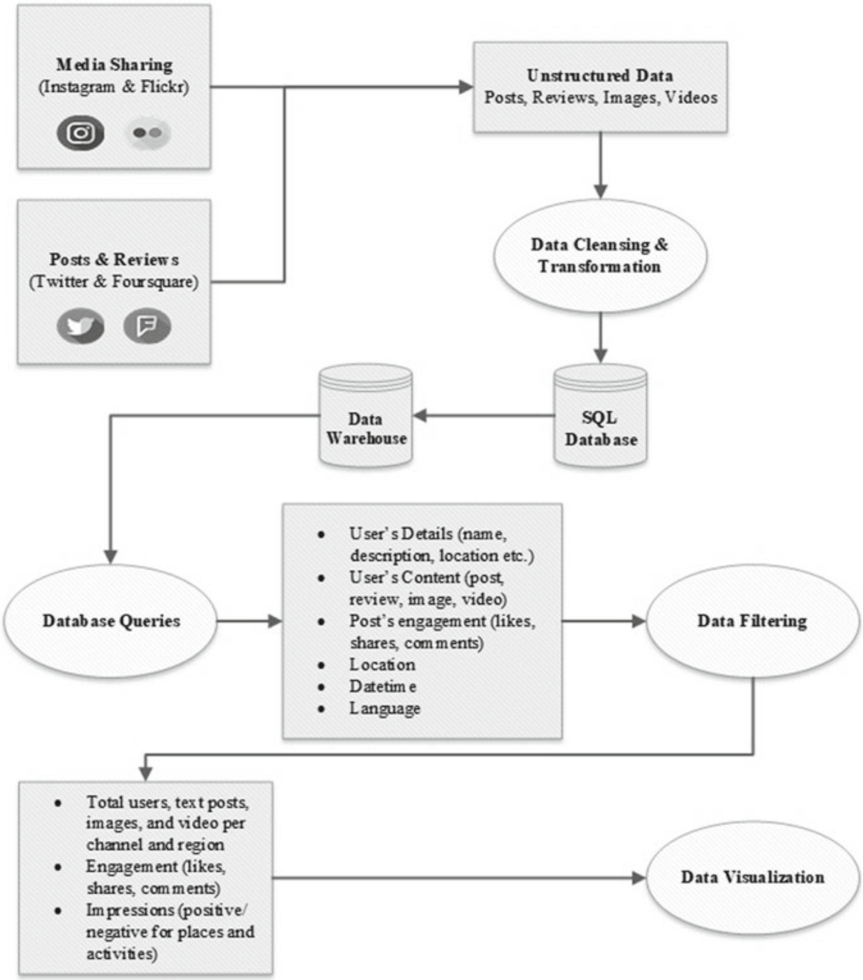


Fig. 7.1 The main data collection analysis stages for tourism social media data

and Instagram more than Flickr for photos. Therefore, it seems that most visitors to both cities used Twitter and Instagram networks for sharing their experiences in social media.

Influencer marketing or e-word of mouth marketing can contribute to enhance destination attractiveness or destination branding since influencers can spread messages affecting communities in the digital world. As a result, it is significant for a brand to collaborate with digital influencers to gain authentic and credible presence in online communities (Uzunoğlu and Kip 2014). As shown in Fig. 7.5 on the 20th of November 2017, Heraklion had an unusually high level of social engagement (likes, shares, and comments). This derived from three Instagram photo uploads by a popu-

Fig. 7.2 Total users, posts, sentiment, photos and engagement for two cities in Crete

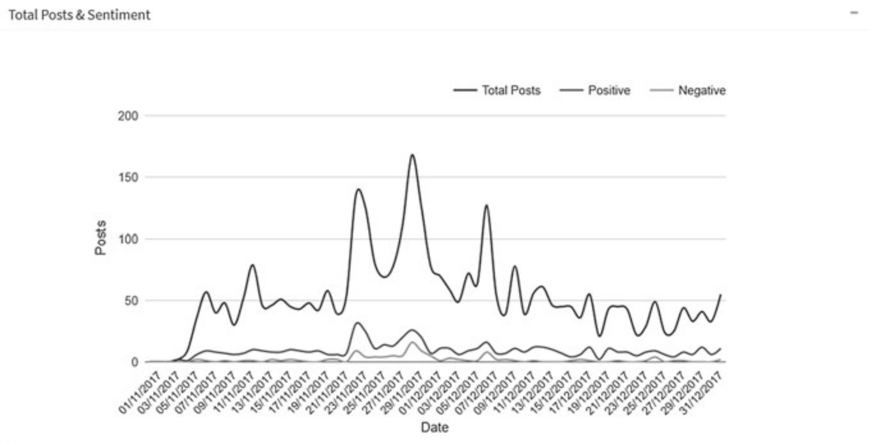


Fig. 7.3 Total posts and sentiment about two cities in Crete

lar singer with 2.3 million followers during her new song's video shooting. Thus, an influencer's posts could be a significant factor for a destination's brand awareness, since influencer marketing is considered to be a viral marketing technique contributing towards building brand awareness (Ferguson 2008).

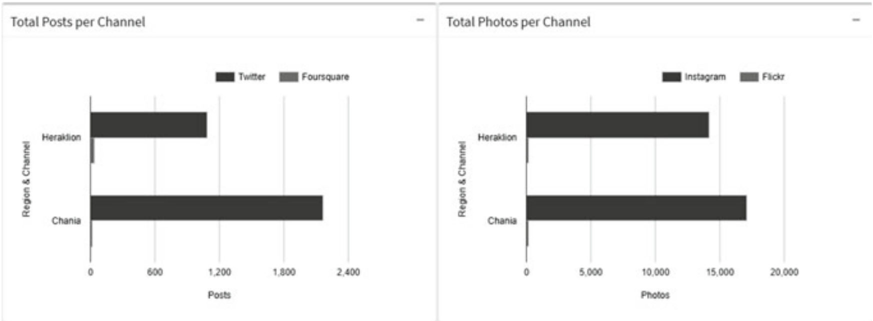


Fig. 7.4 Total posts and photos per channel for two cities in Crete

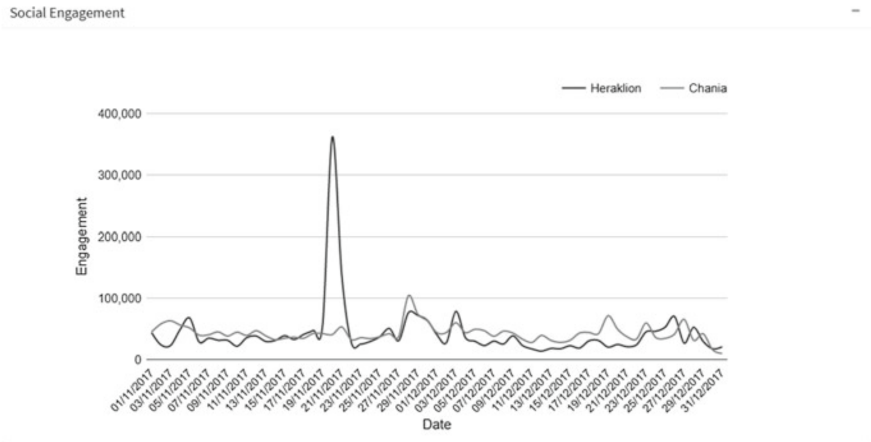


Fig. 7.5 Social engagement for two cities in Crete

7.5 Conclusions

By gathering and analyzing social media data, stakeholders can gain valuable knowledge about tourists' and travelers' preferences and needs. The case study analyses public data from four LBSN platforms (Twitter, Foursquare, Instagram, and Flickr). The Crete results provide knowledge about trends in places among dates, visitor sentiment (highest for traditional food and beautiful places - worst about the quality of services) and users engagement across destinations.

References

Bouadi T et al (2017) A data warehouse to explore multidimensional simulated data from a spatially distributed agro-hydrological model to improve catchment nitrogen management. Environ Model Softw 97:229–242

Brandt T, Bendler J, Neumann D (2017) Social media analytics and value creation in urban smart tourism ecosystems. Inf Manag 54(6):703–713

Carvalho JP et al (2017) MISNIS: An intelligent platform for twitter topic mining. Expert Syst Appl 89:374–388

Chang Y-C, Ku C-H, Chen CH (2017) Social media analytics: extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. Int J Info Manag. Available at http://dx.doi.org/10.1016/j.ijinfomgt.2017.11.001

Chen H, Chiang RHL, Storey VC (2012) Business intelligence and analytics: from big data to big impact. MIS Q Manag Inf Syst 36(4):1165–1188

Chen M, Mao S, Liu Y (2014) Big data: a survey. Mob Netw Appl 19(2):171–209

Chorley MJ, Whitaker RM, Allen SM (2015) Personality and location-based social networks. Comput Hum Behav 46:45–56

Donaire JA, Camprubí R, Galí N (2014) Tourist clusters from Flickr travel photography. Tourism Manag Perspect 11:26–33

Ferguson R (2008) Word of mouth and viral marketing: taking the temperature of the hottest trends in marketing. J Consum Mark 25(3):179–182

Ferreira N et al (2013) Visual exploration of big spatio-temporal urban data: a study of New York City Taxi Trips. IEEE Trans Visual Comput Graph 19(12):2149–2158

Floris R, Campagna M (2014) Social media data in tourism planning: analysing tourists' satisfaction in space and time. In Schrenk et al M (eds) REAL CORP 2014—PLAN IT SMART! Clever solutions for smart cities. 19th international conference on urban planning and regional development in the information society. CORP—Competence Center of Urban and Regional Planning, pp 997–1003

Günther WA et al (2017) Debating big data: a literature review on realizing value from big data. J Strateg Inf Syst 26(3):191–209

Hashem IAT et al (2015) The rise of "big data" on cloud computing: review and open research issues. Inf Syst 47:98–115

Huang A, Gallegos L, Lerman K (2017) Travel analytics: understanding how destination choice and business clusters are connected based on social media data. Transp Res Part C Emerg Technol 77:245–256

Komorowski M, Do Huu T, Deligiannis N (2018) Twitter data analysis for studying communities of practice in the media industry. Telematics Inform 35(1):195–212

Lee R, Wakamiya S, Sumiya K (2011) Discovery of unusual regional social activities using geotagged microblogs. World Wide Web J Biol 14(4):321–349

Li D, Zhou X, Wang M (2018) Analyzing and visualizing the spatial interactions between tourists and locals: a Flickr study in ten US cities. Cities. Available at http://dx.doi.org/10.1016/j.cities.2017.12.012

Liu Y et al (2016) ELAN: an efficient location-aware analytics system. Big Data Res 5:16–21

Liu H et al (2018) Detecting global and local topics via mining twitter data. Neurocomputing 273:120–132

Mahmud S, Iqbal R, Doctor F (2016) Cloud enabled data analytics and visualization framework for health-shocks prediction. Future Gener Comput Syst FGCS 65:169–181

Marchiori E, Cantoni L (2015) The role of prior experience in the perception of a tourism destination in user-generated content. J Destination Market Manag 4(3):194–201

Marine-Roig E, Clavé SA (2015) Tourism analytics with massive user-generated content: a case study of Barcelona. J Destination Market Manag 4(3):162–172

Miah SJ et al (2017) A Big Data analytics method for tourist behaviour analysis. Inf Manag 54(6):771–785

Milton S (2011) Location intelligence—the future looks bright. Forbes. Available at https://www.forbes.com/sites/stevemilton/2011/10/11/location-intelligence-the-future-looks-bright. Accessed 18 Jan 2018

Mittal V et al (2017) Multivariate features based instagram post analysis to enrich user experience. Procedia Comput Sci 122:138–145

Park SB, Jang J, Ok CM (2016) Analyzing Twitter to explore perceptions of Asian restaurants. J Hospitality Tourism Technol 7(4):405–422

Rathore MM et al (2017) Big data analytics of geosocial media for planning and real-time decisions. In: 2017 IEEE international conference on communications (ICC). Available at http://dx.doi.org/10.1109/icc.2017.7996545

Ravi K et al (2018) Analytics in/for cloud-an interdependence: a review. J Netw Comput Appl 102:17–37

Rehman NU, Weiler A, Scholl MH (2014) OLAPing Social Media: the case of Twitter. In: 2013 IEEE/ACM international conference on advances in social networks analysis and mining. Knowledge Discovery and Data Mining. ACM, pp 1139–1146

Reinsel D, Gantz J, Rydning J (2017) Data Age 2025: the evolution of data to life-critical. IDC

Schmidbauer H, Rösch A, Stieler F (2018) The 2016 US presidential election and media on Instagram: Who was in the lead? Comput Hum Behav 81:148–160

Shekhar S, Xiong H (2007) Encyclopedia of GIS. Springer Science & Business Media

Uzunoğlu E, Kip SM (2014) Brand communication through digital influencers: leveraging blogger engagement. Int J Inf Manage 34(5):592–602

Vashisht P, Gupta V (2015) Big data analytics techniques: a survey. In: 2015 international conference on green computing and internet of things (ICGCIoT). Available at http://dx.doi.org/10.1109/icgciot.2015.7380470

Vassakis K, Petrakis E, Kopanakis I (2017) Big data analytics: applications, prospects and challenges. In: Lecture notes on data engineering and communications technologies, pp 3–20

Vecchio PD et al (2017) Creating value from social big data: implications for smart tourism destinations. Inf Process Manag. Available at http://dx.doi.org/10.1016/j.ipm.2017.10.006

Wang L et al (2015) On the brink: predicting business failure with mobile location-based checkins. Decis Support Syst 76:3–13

Yoo K-H, Sigala M, Gretzel U (2016) Exploring TripAdvisor. In: Tourism on the verge, pp 239–255

Zheng Y (2011) Location-based social networks: users. In: Computing with spatial trajectories, pp 243–276

Zhou X, Xu C, Kimmons B (2015) Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. Comput Environ Urban Syst 54:144–153

Chapter 8
Identifying Innovative Idea Proposals with Topic Models—A Case Study from SPA Tourism

Gabriele Sottocornola, Fabio Stella, Panagiotis Symeonidis, Markus Zanker, Ines Krajger, Rita Faullant and Erich Schwarz

Abstract This chapter builds on a dataset where online users of a spa platform participated in an online idea contest providing free-text descriptions of their proposals for spa services. A panel of domain experts annotated these idea descriptions with a score for their innovativeness that serves as ground truth for machine learning experiments. Thus, the contribution lies in the application of topic modeling techniques to free-text idea descriptions in order to automatically identify innovative proposals based on advanced text processing and machine learning. Results of this case study indicate that topic modeling can outperform the ZeroR baseline as well as traditional survey scales for lead user identification and therefore constitute a first step towards exploring this technique for innovation research.

Keywords Spa tourism · Topic modeling · Online idea contest

8.1 Introduction

For services in general, and in particular for tourism offers, the customer has a crucial and integral role. Services are dependent upon the interaction quality between service provider and customer (Bolton and Drew 1991; Grönroos 1993; Parasuraman et al. 1988). Despite the acknowledgement of the customers' central role for

---

This chapter extends the work published in Faullant et al. (2012).

G. Sottocornola (✉) · P. Symeonidis · M. Zanker
Free University of Bozen-Bolzano, Bolzano, Italy
e-mail: gsottocornola@unibz.it

F. Stella
University of Milan-Bicocca, Milan, Italy

I. Krajger · E. Schwarz
Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria

R. Faullant
Syddansk Universitet, Odense, Denmark

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_8

tourism, it is surprising that customers are rarely integrated into new service development for tourism. In contrast, customer integration into new product development established on a systematic basis in the early 1980s for physical goods (Urban and Von Hippel 1988; Von Hippel 1978). In particular, the most demanding and advanced customers, so called lead users (Von Hippel 1986), have been integrated into new product development. Several studies suggest that the integration of users into new product development is an appropriate means for companies to come up with faster and more customer-centered innovations, e.g. (Gruner and Homburg 2000). Also for the service industry the potential for various stages and modes of user involvement has been highlighted (Alam 2002). With our study we test whether the concept of lead-user integration is in principle also applicable for new service development in tourism. Users of an online platform focusing on spa vacations have been invited to participate in a virtual idea competition and to submit their ideas for new service offerings in spa tourism. In addition, participants' lead user characteristics were collected via a web-based questionnaire. Based on creativity theory an independent jury of spa experts evaluated these ideas, which were then correlated with the users' lead user characteristics.

This chapter extends the work presented in Faullant et al. (2012) to automatically identify innovative ideas from users. Our proposed method consists of applying topic modeling, namely Latent Dirichlet Allocation (LDA) (Blei 2012), to represent submitted ideas with their underlying topics. A topic can be seen as a sorted list of words that share a coherent semantic meaning and that are automatically extracted from corpora of natural language documents in an unsupervised manner. A panel of judges that scored each submitted proposal to an idea contest serves as ground truth. The hypothesis of this work is that the topical representation of each document can be effective to discern which formulated ideas (and thus also which participants) are innovative in the context of this study with respect to spa tourism.

In order to classify which of the participants' submissions better fit the innovation and customer's value expected by the judges we trained a tree-based classification model. The main advantages are that, once the model is trained, we can automatically have a degree of confidence on the innovation of a submitted idea together with an indication of the most important features, i.e. the topics, for the classification model. Since LDA provides an effective way to represent topics through their most representative words, we can try to explain which topics, and thus terms, can support the classification of an idea as innovative by the panel of judges.

The case study addresses the following research questions:

- Is LDA an effective technique to identify innovative traits or features within textual descriptions of ideas?
- Are tree-based classifiers together with LDA suitable to understand why a submitted idea is classified as innovative or not?

The rest of the paper is organized as follows: Sect. 8.2 presents the literature review of the field of lead users integration in new products development, together with the methodologies involved in this case study and some related works in the field of e-tourism; Sect. 8.3 describes the experiments conducted on the specific dataset of spa tourism; Sect. 8.4 presents the results and finally conclusions are drawn.

8.2 Related Works

This chapter describes the application of text mining and machine learning techniques for the identification of opportunities for new product and service development. Thus, the related work covers all these aforementioned aspects. In addition, the subsections on LDA and classification provide a tutorial-like introduction.

8.2.1 Lead Users in New Product Development

Within the literature on innovation management, user integration into new product development has become an important research field. Instead of solely considering users as information providers, users can actively engage in the new product development process (Edvardsson et al. 2012; Von Hippel 2005; Campos et al. 2018). Previous research confirms the ability of users to contribute to the NPD (new product development) process (Alam 2006; Füller et al. 2007; Lilien et al. 2002; Oliveira and von Hippel 2011; Skiba and Herstatt 2009). However, only a small proportion of users, between 10 and 40%, has the know-how, creativity and expertise for truly innovative problem solutions that are not restricted to simple extensions or incremental innovations (Von Hippel 2005). Since the value of customer contributions to the development of new products and services varies significantly, it is crucial to carefully select the right users to be integrated into new product or service development (Enkel et al. 2005; Gruner and Homburg 2000; Wellner 2015). One group of users that has been shown to be able to deliver highly innovative suggestions for new product development are lead users. Lead Users are different from other users because they (a) have needs that will become commonplace in a market before the bulk of the other users encounters them and (b) they expect to benefit significantly from obtaining a solution to those needs (Von Hippel 1986). These characteristics are also known as the "Ahead of Trend" (AT) dimension and the "High Expected Benefit" (HEB) dimension respectively (Franke et al. 2006). Products developed in cooperation with lead users are appreciated as highly innovative by firms (Franke et al. 2006; Lilien et al. 2002; Lüthje 2000). The ability to bear innovative solutions is fundamentally linked to a person's individual creativity (Faullant et al. 2009). In psychology, creativity is generally defined as "the production of novel, useful ideas or problem solutions" (Amabile et al. 2005). The first aspect emphasizes the originality or unexpectedness of an idea (Sternberg and Lubart 1999). The second aspect stresses that an idea must be of value, or "appropriate" (i.e., useful, adaptive concerning task constraints) (Sternberg and Lubart 1999) which is especially important for new product development. Many studies have confirmed that lead users are able to produce both novel and useful ideas.

Initial lead user studies concentrated predominantly on the industrial goods markets (Franke and Von Hippel 2003; Herstatt and Von Hippel, 1992; Lüthje 2003; Morrison et al. 2000; Olson and Bakke 2001; Urban and Von Hippel 1988). The identification of lead users is also promising for user integration in consumer mass markets such as kite surfing, extreme sporting equipment, technical diving, and kitchen appliances (Franke and Shah 2003; Füller et al. 2006, 2007; Lüthje 2004; Lüthje et al. 2005; Schwarz et al. 2009). Within new service development, systematic lead user identification and their integration for service innovation has been widely neglected (Skiba and Herstatt 2009). Edvardson et al. (2012) proposed a conceptual framework of methods of customer integration into new service development. The lead user method was characterized as being able to generate highly novel service solutions, but at the same time requiring high methodological competences. Recent empirical evidence confirms the potential of user innovation for the service sector (Oliveira and von Hippel 2011). In tourism so far, little is known about lead user identification and their involvement in new service development.

8.2.2 Virtual User Integration for New Product and Service Development

The use of the Web allows companies to reach potential users world-wide for new product development (Füller and Hienerth 2004; Sawhney et al. 2005). This is accompanied by the development of new tools and methods for virtual user integration, e.g. (Dahan and Hauser 2002; Dahan and Srinivasan 2000; Franke and Piller 2004; Füller et al. 2007; Jeppesen 2005; Verona et al. 2006). Web-based methods such as idea competitions, toolkits for user innovation, virtual worlds, virtual stock markets and virtual communities have already diffused in practice supporting collaborative new product development (Bullinger et al. 2010; Ebner et al. 2009). For the service sector in general Sigala (2010) provided insights from the Starbucks community that virtual user communities are able to generate, shape, and co-create ideas for new service development. The shared interpretation of an idea throughout the community can lead to different cultural interpretations of what a new service might constitute. Another study within the mobile service industry also demonstrates the potential of a firm hosted virtual lead user community for new service development (Mahr and Lievens 2012). In tourism, the potential of user communities for the development of new tourism products was recognized in the early 2000s (Wang et al. 2002). Recent studies confirm the importance of customer co-creation in travel services and its impact on customer satisfaction and expenditure level (Grissemann and Stokburger-Sauer 2012). These findings advocate against findings from earlier studies that demonstrate that user activities in user communities and blogs are still limited to information exchange, such as sharing and documenting travel experiences and ratings of tourism products, more recently an active role of users in co-creation of tourism value creation has been acknowledged (Dippelreiter et al. 2008; Waldhör and Rind 2008; Yoo and Gretzel 2008; Rihova et al. 2015). Especially through the use of technology users can actively engage in shaping their tourism experience (Neuhofer et al. 2014). Meanwhile a range of successful examples of user involvement and user co-creation in the hospitality and leisure sector have been demonstrated (Egger et al. 2016). With our study we investigate whether users have the potential to substantially contribute to new service development and whether those users can be identified by web-based means.

8.2.3 Latent Dirichlet Allocation and Topic Models

Topic models are a suite of algorithms that aim to discover the main themes, denoted as topics, that pervade a large and otherwise unstructured collection of natural language documents. Topic models are able to annotate and summarize this corpus with the thematic information provided by topics.

The main contribution to probabilistic topic modeling was provided by Blei (2012) with the introduction of *Latent Dirichlet Allocation (LDA)*, where the underlying idea is that each document contains a mixture of multiple topics. LDA can be easily described through its *generative process*, a simple probabilistic procedure by which documents can be ideally generated.

Consider the visual representation of the LDA generative process for a corpus of tourism-related reviews, represented in Fig. 8.1. A topic is formally defined as a probability distribution over a fixed vocabulary (the corpus dictionary) and a review document is associated with a probability distribution over topics. Assuming that the distribution of the topics over words is given (boxes on the left), the generative process of a generic document d in the corpus consists of the following steps:

1. Randomly choose a distribution over topics for document d (histograms on the right);
2. For each word in the document d:
   a. Randomly choose a topic z from the distribution over topics sampled at step 1 (coins);
   b. Randomly choose a word w from the corresponding distribution over the vocabulary, given the sampled topic z (word assignment in the text).

More formally the process specifies the probability of sampling a specific word token w_i as follows:

$$P(w_i) = \sum_{j=1}^{T} P(w_i|z_i = j)P(z_i = j)$$

where $P(z_i = j)$ is the probability that j-th topic was assigned to the i-th word token, $P(w_i|z_i = j)$ is the conditional probability to extract i-th word token given the assignment of topic j.

This statistical generative model reflects the intuition that documents contain multiple topics. Each document contains the topics in different proportion (step 1); each word in each document is drawn from one of the topics (step 2b), where the selected topic is chosen from the per-document distribution over topics (step 2a).

The aim of LDA model is to invert this generative process: the occurrences of words in the documents are the observed variables, while the topic structure (i.e. per-word topic distributions and per-document topic distributions) is hidden. By exploiting techniques of statistical inference and sampling (i.e. Gibbs sampling and variational bayesian inference), these probability distributions are inferred by observing the frequency of words within documents.

Some applications of topic models to e-tourism and recommendation systems in general have been described.

Rossetti et al. (2016) provide a description of topic models with a particular focus on the tourism domain. They propose different application scenarios where the topic models effectively processes textual reviews in order to provide decision support and recommendations to online tourists as well as to build a basis for further analytics (i.e. provide additional semantics for explanation and understanding of the enormous amounts of user-generated data). Furthermore, the contribution consists of two new models based on LDA (namely, the *topic-criteria model* and the *topic-sentiment criteria model*) and results from experimenting with user-generated review data on restaurants and hotels.

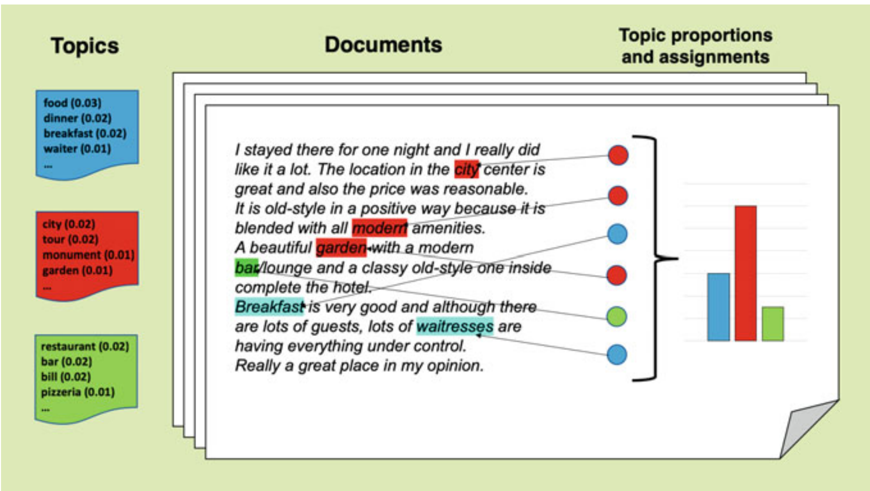


Fig. 8.1 A visual representation of the generative process underlying LDA

Wang et al. (2010) proposed a probabilistic generative model similar to LDA applied to textual reviews on hotels to estimate opinion ratings on topical aspects (e.g. cleanliness, location or sleep quality), a problem defined as *Latent Aspect Rating Analysis (LARA)*. The underlying assumption is that each sentence in a review is related to a specific aspect. The proposed generative model assumes that for each sentence a user decides which aspect she wants to write about and chooses the words to write accordingly. To assign one or more aspects to each sentence a bootstrap procedure is defined in order to provide keyword sentences to different aspects. The method is able to distinguish between cases where the overall ratings are the same but aspect ratings are different. Furthermore, review analysis opens a range of possible applications, such as opinion summarization on topical aspects, ranking of entities based on these aspect ratings and the analysis of the rating behavior of reviewers.

Agarwal and Chen (2010) introduced a matrix factorization method for recommender systems where items have a natural bag-of-word representation named *fLDA*. The method works by regularizing both user and item factors simultaneously through user features and the bag of words associated with each item. In particular, each word in an item is associated with a discrete latent factor (i.e. the topic of the word); item topics are obtained by averaging topics across all words in an item. Then, a user rating on an item is modeled as user's affinity to the item's topics where users' affinity to topics (i.e. user factors) and topic assignments to words in items (i.e. item factors) are learned jointly in a supervised fashion. Topics extracted from item descriptions and user metadata are exploited as priors to regularize item and user latent factors. The posterior distribution of item and user factors depends on both the prior and user ratings on items, since the LDA model is exploited to regularize item latent factors, and the Gaussian linear regression regularizes user latent factors. The model has been proven to be accurate and capable to deal with warm-start and cold-start scenarios, as textual data related to new users and new items can be used to compute recommendations. Furthermore, it provides interpretable latent factors that can explain user-item interactions.

McAuley and Leskovec (2013) aimed to combine latent rating dimensions (i.e. latent-factor recommender systems) with latent review topics (i.e. LDA topic models) in order to estimate the ratings from textual reviews on different datasets. The *Hidden Factors as Topics (HFT)* approach consists of two steps: first, latent factors for rating prediction are fitted, and second, topic assignments to item reviews are updated merging item-topic distributions to its latent factors. The proposed approach not only leads to more accurate predictions on recommendations, but can also solve side problems. First, it deals with the cold-start problem, exploiting content topics for items with only a few ratings. Second, it is able to discover and automatically categorize items in different categories based on the topics discussed in the reviews. Third, it is able to identify representative reviews, which can be shown to users as an explanation of item characteristics.

Another extension of LDA applied to user reviews is the *Joint Sentiment-Topic model (JST)* (Lin and He 2009). In contrast to the majority of sentiment analysis models which are based on classification models, this model is able to extract sentiment and topics simultaneously from text in an unsupervised way. The main difference with respect to the LDA model is that JST adds an additional sentiment layer between the document and the topic layer. In this way a four level hierarchy is defined where documents have distributions on sentiment labels, sentiment labels have distributions on topics and topics have distributions on words. The model has been evaluated on the movie review dataset to classify the review sentiment polarity to further improve the sentiment classification accuracy.

8.2.4 Decision Tree Based Classification Models

Decision tree classification models (also referred to as decision tree learning) are a family of predictive modeling approaches used for supervised classification in the field of machine learning (Rokach and Maimon 2014). A decision tree is used as a predictive model to go from observations of the features of an item (represented in the branches) to conclusions about the item's target class (represented by the leaves). Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.

A decision tree can be learned by splitting the data into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the whole subset belongs to the same target class, or when splitting no longer adds value for the classification. Decision tree learning is the construction of a decision tree from class-labeled training tuples. A decision tree is a flow-chart-like structure, where each internal (non-leaf) node denotes a test on an attribute, each branch represents the outcome of a test, and each leaf (or terminal) node holds a class label. The topmost node in a tree is the root node. Many specific learning algorithms were proposed in the literature throughout the years, and the most relevant are: ID3, C4.5, and CART.

Algorithms for constructing decision trees usually work top-down, by choosing a variable at each step that best splits the set of items. Different algorithms use different metrics for measuring the best split. These generally measure the homogeneity of the target variable within the subsets. These metrics are applied to each candidate subset, and the resulting values are combined (e.g. averaged) to provide a measure of the quality of the split. The most used metrics to determine the quality of a split are Gini impurity and information gain.

A particular type of decision tree classifier that had lot of success in the machine learning community is the random forests classifier. *Random forests* were first introduced by Breiman (2001), which describes a method of building a forest of uncorrelated trees, combined with randomized node optimization and bagging. Random forests are an ensemble learning method for classification, that operates by constructing a multitude of decision trees at training time and outputs the class that is the mode of the predicted classes of the individual decision trees. The main advantage of random forests over decision trees is the capability of avoiding overfitting to the training

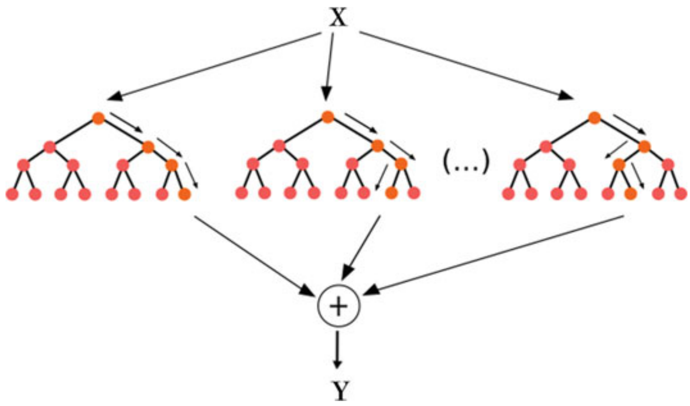


Fig. 8.2 A graphical representation of a Random forest

set, thanks to the bagging (or bootstrap aggregating) procedure. Figure 8.2 provides a visual example of a random forest structure.

8.3 Methodology

In this section we describe the experimental framework to evaluate the effectiveness of the LDA methodology to identify innovative users with a case study from the spa domain. We first illustrate the process of data collection and the dataset characteristics. Then we describe in details the preprocessing process and the experiments performed on data. Finally, we present and discuss the results.

8.3.1 Dataset

For evaluating the proposed approach the dataset presented in (Faullant et al. 2012) was used. The data was collected through an empirical study executed in the field of a search engine to retrieve curated information on European spas. Visitors of the website were invited to submit innovative ideas for spa development and new service creation. After describing the details of their idea in a free text field, participants were asked to complete a standardized web-based questionnaire. To assess participants' for being a lead user, the two hallmark characteristics of being *ahead of trend* (AT) and *high expected benefit* (HEB) were used from existing measures in the literature (Lüthje 2000). The scales were adapted to the spa context and were measured as a continuous variable on a 7-point Likert scale as denoted in Table 8.1. Furthermore,

Table 8.1 Codes of questionnaire items

Code	Scale
AT1	I'm regarded as being well informed in the field of spa offers
AT2	I usually determine new spa offers earlier than most other people
AT3	I try to visit just recently opened spas
HEB1	I have needs and preferences which are not satisfied by spa offers
HEB2	During my past visits of spa resorts I noticed shortcomings several times
HEB3	I'm dissatisfied with the existing spa resort offers
SPA1	How often do you visit spa resorts each year?
SPA2	How many different spa resorts have you visited up to now?
SPA3	Which of the following eight recently opened spa resorts have you visited?

demographic data like gender, age, nationality and education as well as information about the actual spa usage of participants were collected.

To evaluate and rank the submitted ideas the *Consensual Assessment Technique* (CAT) (Amabile 1982) was applied. According to this method "a product or response is creative to the extent that appropriate observers independently agree it is creative." (Amabile 1982). The quality of the submitted ideas was independently evaluated by a jury of 4 experts in the spa domain, based on 3 dimensions: originality of the idea (*orig*), customer value of the idea (*util*), overall impression (*over*). These dimensions were presented on a scale from 0 (no value) to 5 (very high value). The experts rated all ideas independently from each other.

In total 161 participants filled out the questionnaire, and submitted 122 ideas or suggestions for spa service development. The dataset contains only users which have submitted an idea (more than 1 term) and which have completed the questionnaire. For 6 users, a missing answer in the questionnaire was replaced with the average value for that question. Finally, after data cleaning the dataset resulted in 116 instances.

Descriptive statistics of replies to the adapted questionnaire on AT and HEB items (scale 1–7, 1: I do not agree at all, 7: I fully agree) are given in Table 8.2. Before preprocessing the textual data, submitted ideas have an average length of 555 characters (standard deviation: 1411.5, min: 17, max: 14501) and Fig. 8.3 depicts their length distribution with one extremely long outlier. The most dominant themes for innovation are the different needs of adults and children that should be addressed in separate locations and ideas for designing the relaxation and recreation areas. On average respondents have 7.18 spa visits per year (SPA1), know 6.25 different spa resorts (SPA2) and have already tried 1.4 out of the 8 newly opened spa resorts that were named in the questionnaire (SPA3).

Table 8.2 Statistics on questionnaire items (Cronbach alpha = .74)

Code	Mean	Std. Dev.
AT1	4.03	1.87
AT2	4.4	1.89
AT3	4.36	1.78
HEB1	2.53	1.68
HEB2	4.44	1.69
HEB3	2.72	1.66

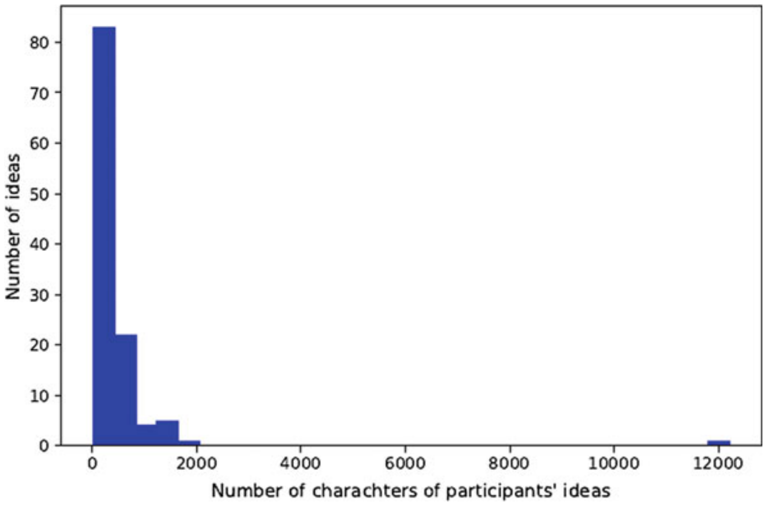


Fig. 8.3 Distribution of character length of the submitted ideas

8.3.2 Preprocessing

We preprocessed the raw data before applying the topic model and the classification algorithm. First, we cast the problem to a binary classification problem, thus, for each one of the previously defined dimensions we compute the majority vote on the judges ratings. Formally, given the idea submitted by user i and a quality dimension d, the class $C_i^d = 1$ if at least 3 out of 4 judges have rated the idea greater or equal to 3 for the dimension d, otherwise $C_i^d = 0$.

We applied transformation to the corpus of users' submitted ideas. Textual content was tokenized, considering non-alphanumeric characters as separators, and all the extracted tokens were converted to lowercase. Furthermore, tokens with less than 3 characters were filtered out together with common German stopwords. Stemming wasn't applied to the tokens to avoid merging term roots with different meanings. The corpus is in German and the dictionary after this process contains 2449 unique German terms.

Table 8.3 Summary of the statistics related to the positive and negative classes for the *over* dimension, i.e. number of users (*users*), total number of characters (*chars*), total number of terms (*terms*), and number of unique terms (*unique*) in the submitted ideas

	Users	Chars	Terms	Unique
Positive	33	24,699	3422	1752
Negative	83	14,625	1984	1100
Total	116	39,324	5406	2449

The class label for our dataset is judges' overall impression (*over*), since it summarizes the other 2 dimensions. In Table 8.3 we report some statistics about the spa dataset with respect to positive and negative instances of this dimension.

Note the following observations from Table 8.3: The number of instances in our dataset is relatively small, thus, we have to expect that a machine learning approach should perform poorly in terms of accuracy and confidence of the predictions. The distribution of the two classes is unbalanced, with more than 70% of the observations being negative. Thus, it should be more difficult for a classifier to get good predictive performance on the minority class (i.e., the positive one). Despite the fact that the majority of the observations are negative, we can see how the number of characters is almost doubled for positive ideas (average length 748 characters) with respect to the negative ones (average length 176 characters). The same observations can be drawn from the counts of the number of terms, i.e. the average number of terms for a positive idea is 104, while the average number of terms for a negative one is 24. Finally, despite the big difference in the length of the ideas between the two classes, there's less difference in the number of unique terms.

8.3.3 Experimental Setup

We designed an evaluation protocol to cope with the very low number of instances to train and test our classification approach. First, we applied a 5-fold cross validation (CV). The data were sampled with a stratified sampling technique to maintain a similar distribution of the target class in every folder. At each CV loop, we learned different LDA topic models on the users' submitted ideas that belong to the training data. We trained LDA with number of topics $T = 3, 5, 7, 10, 13, 15, 20, 30$. To compute these models, we used the Python wrapper for the well-known NLP library MALLET.[1] The library implements an effective and efficient version of LDA, based on Gibbs sampling. We set the number of learning iterations to 1000 with hyperparameter optimization every 10 iterations (after 200 iterations of burn-in). The trained models are further exploited to infer the doc-topic distributions (i.e., the proportion

[1]McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.

of the topics hidden in each document) for the test corpus of ideas. We held-out test data from the training of the LDA model, in order to not overfit the model to that corpus, even with an unsupervised technique. After this process we came out with a new set of features (namely, the doc-topic distribution for each textual idea) for each trained topic model. These features were further exploited for the classification task.

For sake of clarity we wanted to narrow down the number of candidate topic models to learn. In literature several methods are provided, e.g. based on perplexity of hold-out documents (Wallach et al. 2009) or coherence scores related to pointwise mutual information (Newman et al. 2010). Since our dataset is too small, we didn't have enough data to provide a significant generalization on this kind of unsupervised measures. Thus, we conducted a preliminary analysis on the whole set of topic models: we trained a separate 10-fold cross validation and we selected the optimal models, i.e. the ones which performed well in terms of accuracy on the prediction task. At the end of this process we manually set the topic models parameter to $T = 5, 10, 15$.

We used each feature set generated by the LDA models to separately train a supervised classification model to predict the target class of interest. At every CV loop, the performances of all the models were evaluated on the same test folder, never seen by the training algorithm, in terms of accuracy, and precision-recall on the positive class. We decided to apply a tree-based classifier, for the advantage of automatically assessing the importance of each feature, based on the computed Gini index (Breiman 2001). This characteristic provides a sort of explanation on which are the specific topics (thus, the most representative words for that topics) mainly contribute to the classification accuracy, to discern between innovative and not innovative ideas.

For our experiments we used the Random Forest classifier provided by Python library scikit-learn.[2] We trained the algorithm with 10 tree estimators with Gini index as a splitting criterion and all the features considered at each splitting point. We employed the same algorithm on the same training data points to build different baseline models to benchmark the classification performance of the topic models. Specifically, we trained two different kind of models: one on the features related to the self-reported characteristics of users and another on the features related to the length of the submitted idea (i.e. the number of words and the number of characters). Furthermore, we merged the two sets of features with the ones extracted by the topic models in order to train hybridized models that are able to capture the information provided by different sources.

8.4 Results and Discussion

In this section we present and discuss the results achieved by our topic model classification approach on the *over* dimension. The results are presented together with

[2]http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.breiman:2001Classifier.html.

three baselines: the classification model trained on the features related to the text length (i.e. *Length*), the classification model trained on the features extracted from the questionnaire (i.e., *Survey*), and the *ZeroR* classification model (i.e. the most trivial predictor that classifies every instance with the most frequent class).

Hence we present the results of the experiments related to the classification performance of the hybrid models, namely, the ones that are built upon the features of each candidate topic model merged together with the features of the two baseline models (i.e., length-related features and survey-related features).

Finally we provide to the reader an insight into the effectiveness of explaining the predictions provided by the proposed technique.

For each trained model we evaluate its performance with different classification metrics, directly derived from the confusion matrix. The confusion matrix is a table with two rows and two columns, that reports the number of *true positives* (TP, i.e. the number of instances that belong to the positive class predicted as positive), *false positive* (FP, i.e. the number of instances that belong to the positive class predicted as negative), *true negatives* (TN, i.e. the number of instances that belong to the negative class predicted as negative) and *false negatives* (FN, i.e. the number of instances that belong to the negative class predicted as positive), collected after the evaluation of a binary classifier. With these values we are able to compute the following metrics, that well describe the effectiveness of our classification algorithms:

- *Precision*, defined as $TP/(TP + FP)$.
- *Recall*, defined as $TP/(TP + FN)$.
- *Accuracy*, defined as $TP + TN/(TP + FP + TN + FN)$.

In Table 8.4 we report the results achieved by the classifiers built upon the candidate topic models (i.e. $T = 5$, $T = 10$, $T = 15$) and the baseline methods (*Length*, *Survey*, *ZeroR*) for the dimension $d = over$.

First, increasing the number of topics does not automatically lead to better performances; the best performing topic models have 5 and 10 topics. While the accuracy is almost the same (with a maximum variation of 1%) for the three models, recall and precision on the positive class deviate significantly from the two best models to the one with 15 topics, i.e. there's a 13% improvement in the precision and more than 5% improvement in the recall. Again, we can see how each of the topic-related classification model strongly outperforms at least two of the baselines in every metric. In particular the *Survey* baseline, i.e. the self-reported characteristics w.r.t. the innovativeness, behave really poorly with respect to the other methods, confirming

Table 8.4 Classification performances of the topic models (i.e. $T = 5$, $T = 10$, $T = 15$) and the baselines (*Length*, *Survey*, *ZeroR*) for the dimension $d = over$

	T = 5	T = 10	T = 15	Length	Survey	ZeroR
Precision	.533	.533	.4	.564	.004	0
Recall	.186	.152	.1	**.448**	.029	0
Accuracy	.734	.742	.743	.748	.614	.716

the findings in (Faullant et al. 2012). Furthermore, from the results it emerges that the classification model based only on the length of the submitted idea is able to closely outperform even the proposed topic model approach, especially for the recall metric (28% improvement).

We take advantage of this finding and therefore explore the performance of our approach when we hybridize the features extracted by the LDA with the other two kinds of features, *Survey* and *length*.

In Table 8.5 we report the results achieved on the classification task by the hybrid models, constructed from the union of the features of the candidate topic models (i.e. $T = 5$, $T = 10$, $T = 15$) with the baselines' features of Length (L) and Survey (S) for the dimension $d = over$.

The results clearly highlight that the hybridization process improves the performance of each topic model configuration, in particular *Length* features in combinations with the topics are able to achieve optimal results. Each hybrid model built with length-based features outperform the results achieved by the topic models and the *Length* model in precision and accuracy. A significant case is represented by the 10-topics model, in which the combination with *Length* allows to achieve the best overall performance for all metrics (except that recall is still slightly outperformed by the simple length-based method). Even the *Survey* features in combination with the 10 topics are able to improve the model with respect to the separated ones. Finally, the noisy *Survey* features overwhelm the model with less topics, $T = 5$, with a significant decrease in accuracy and precision.

Finally, we provide an insight of the advantage of the proposed approach with respect to the *explainability* of the classification outcome. As stated above in this Section, a tree-based classifier is naturally able to rank features with respect to their importance in the classification process (i.e., the importance of a feature is computed as the Gini importance, which roughly represents the percentage of instances that a particular feature contributes to correct classification). Furthermore, each feature in our topic model approach can be represented as a ranked list of words, since LDA automatically extract a set of topics, which are a probability distribution over the vocabulary. The combination of the two models can provide an hint, that is easy to understand, on which are the clusters of words (i.e., the topics) that most effectively distinguish between innovative and not innovative ideas.

Table 8.5 Classification performances of the hybrid models that combine the features of the candidate topic models (i.e. $T = 5$, $T = 10$, $T = 15$) with the baselines' features of Length (L) and Survey (S) for the dimension $d = over$

	T = 5 + L	T = 5 + S	T = 10 + L	T = 10 + S	T = 15 + L	T = 15 + S
Precision	.603	.267	**.636**	.55	.587	.6
Recall	.391	.114	.423	.162	.338	.095
Accuracy	.759	.706	**.767**	.752	.76	.742

Table 8.6 Some examples of the explanation technique provided by our approach: the feature importance in the classification (*FeatureRel*) is related to a topic, which has a certain predominance in the corpus (*TopicRel*) and is represented by a ranked list of words (*TopWords*)

FeatureRel	TopicRel	TopWords
.32	.345	erding restaurants wasser ruheräume bad hotel moderne aufguss saunen hundebesitzer
.24	.362	außenbereich wasser erding restaurants ruheräume sauna warmes wellnessbereich achten erlebnisaufgüsse
.41	.016	thema vorstellen http www besucher show ideen watch youtube freizeitparks
.42	.043	man thermen idee vorstellen http besucher www thema show youtube

In Table 8.6 we report some illustrative examples of how this technique can positively impact our case study, providing a further explanation on the predictions. In particular, we give 4 examples, taken from the trained 5-topics models, of the most representative words for the most important features.

The examples show that there's not a clear correlation between the importance of a feature, as provided by the tree-based classifier (*FeatureRel*) and the relevance of a topic in a corpus (*TopicRel*, i.e. the percentage of documents that belong to it). The predominance of a topic in the corpus is not enough to indicate its predictive power. Furthermore, the relevant topics (automatically extracted from the corpus of ideas in a unsupervised manner) are the ones that are represented by significant words (*TopWords*), related to a particular type of room or to activity, e.g. ruheräume (relax room), außenbereich (outdoor activities), aufguss (infusions), hundebesitzer (dog owner), wellnessbereich (wellness area), restaurants, freizeitparks (leisure park), or related to particular video content the users want to share with the judges (as in the last two examples presented).

8.5 Conclusions

This case study explored the application of LDA techniques for identifying innovative users based on free-text submission to an online idea contest. The contribution of this chapter lies in presenting the application of this technique in combination with learning classifiers to a wider audience of tourism and innovation researchers. Although sample size is a limitation for this case-study, the achieved results demonstrate that the prediction accuracy of traditional surveys for lead user identification as well as a ZeroR baseline can be surpassed. Furthermore, results can be even slightly improved if the feature space of identified topics is extended by considering also ideas' length or responses to the survey for lead user identification.

References

Agarwal D, Chen B-C (2010) Flda: matrix factorization through Latent Dirichlet Allocation. In Proceedings of the third ACM international conference on web search and data mining. ACM, pp 91–100

Alam I (2002) An exploratory investigation of user involvement in new service development. J Acad Mark Sci 30(3):250

Alam I (2006) Removing the fuzziness from the fuzzy front-end of service innovations through customer interactions. Ind Mark Manage 35(4):468–480

Amabile TM (1982) Social psychology of creativity: a consensual assessment technique. J Pers Soc Psychol 43(5):997

Amabile TM, Barsade SG, Mueller JS, Staw BM (2005) Affect and creativity at work. Adm Sci Q 50(3):367–403

Blei DM (2012) Probabilistic topic models. Commun ACM 55(4):77–84

Bolton Ruth N, Drew James H (1991) A multistage model of customers' assessments of service quality and value. J Consum Res 17(4):375–384

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Bullinger AC, Neyer A-K, Rass M, Moeslein KM (2010) Community-based innovation contests: where competition meets cooperation. Creativity and Innov Manag 19(3):290–303

Campos AC, Mendes J, do Valle PO, Scott N (2018) Co-creation of tourist experiences: a literature review. Curr Issues Tourism 21(4):369–400

Dahan E, Hauser JR (2002) The virtual customer. J Prod Innov Manag 19(5):332–353

Dahan E, Srinivasan V (2000) The predictive power of internet-based product concept testing using visual depiction and animation. J Prod Innov Manag 17(2):99–109

Dippelreiter B, Grün C, Pöttler M, Seidel I, Berger H, Dittenbach M, Pesenhofer A (2008) Online tourism communities on the path to web 2.0: an evaluation. Inf Technol Tourism 10(4):329–353

Ebner W, Leimeister JM, Krcmar H (2009) Community engineering for innovations: the ideas competition as a method to nurture a virtual community for innovations. R&D Manag 39(4):342–356

Edvardsson B, Kristensson P, Magnusson P, Sundström E (2012) Customer integration within service development: a review of methods and an analysis of insitu and exsitu contributions. Technovation 32(7–8):419–429

Egger R, Gula I, Walcher D (2016) Towards a holistic framework of open tourism. In: Open tourism. Springer, pp 3–16

Enkel E, Perez-Freije J, Gassmann O (2005) Minimizing market risks through customer integration in new product development: learning from bad practice. Creativity Innov Manag 14(4):425–437

Faullant R, Schwarz EJ, Kraiger I, Breitenecker RJ (2009) Towards a comprehensive understanding of lead userness: the role of personality and creativity. In: 16th international product development management conference

Faullant R, Krajger I, Zanker M (2012) Identification of innovative users for new service development in tourism. In: 19th conference on information and communication technologies in tourism (ENTER). Springer, Helsingborg, Sweden

Franke N, Piller F (2004) Value creation by toolkits for user innovation and design: The case of the watch market. J Prod Innov Manag 21(6):401–415

Franke N, Shah S (2003) How communities support innovative activities: an exploration of assistance and sharing among end-users. Res Policy 32(1):157–178

Franke N, Von Hippel E (2003) Satisfying heterogeneous user needs via innovation toolkits: the case of apache security software. Res Policy 32(7):1199–1215

Franke N, Von Hippel E, Schreier M (2006) Finding commercially attractive user innovations:a test of lead-user theory. J Prod Innov Manag 23(4):301–315

Füller J, Hienerth C (2004) Engaging the creative consumer. Eur Bus Forum 19:54–57

Füller J, Bartl M, Ernst H, Mühlbacher H (2006) Community based innovation: how to integrate members of virtual communities into new product development. Electron Commer Res 6(1):57–73

Füller J, Jawecki G, Mühlbacher H (2007) Innovation creation by online basketball communities. J Bus Res 60(1):60–71

Grissemann US, Stokburger-Sauer NE (2012) Customer co-creation of travel services: the role of company support and customer satisfaction with the co-creation performance. Tour Manag 33(6):1483–1492

Grönroos C (1993) Toward a third phase in service quality research: challenges and future directions. Adv Serv Mark Manag 2(1):49–64

Gruner Kjell E, Homburg Christian (2000) Does customer interaction enhance new product success? J Bus Res 49(1):1–14

Herstatt C, Von Hippel E (1992) From experience: developing new product concepts via the lead user method: a case study in a low-tech field. J Prod Innov Manag 9(3):213–221

Jeppesen LB (2005) User toolkits for innovation: consumers support each other. J Prod Innov Manag 22(4):347–362

Lilien GL, Morrison PD, Searls K, Sonnack M, von Hippel E (2002) Performance assessment of the lead user idea-generation process for new product development. Manage Sci 48(8):1042–1059

Lin C, He Y (2009) Joint sentiment/topic model for sentiment analysis. In: Proceedings of the 18th ACM conference on information and knowledge management. ACM, pp. 375–384

Lüthje C (2000) Kundenorientierung im Innovationsprozess. Eine Untersuchung der Kunden-Hersteller-Interaktion in Konsumgütermärkten, Wiesbaden

Lüthje C (2003) Customers as co-inventors: an empirical analysis of the antecedents of customer-driven innovations in the field of medical equipment. In: Proceedings of the 32th EMAC conference, Glasgow

Lüthje C (2004) Characteristics of innovating users in a consumer goods field: an empirical study of sport-related product consumers. Technovation 24(9):683–695

Lüthje C, Herstatt C, Von Hippel E (2005) User-innovators and local information: the case of mountain biking. Res Policy 34(6):951–965

Mahr D, Lievens A (2012) Virtual lead user communities: drivers of knowledge creation for innovation. Res Policy 41(1):167–177

McAuley J, Leskovec J (2013) Hidden factors and hidden topics: understanding rating dimensions with review text. In: Proceedings of the 7th ACM conference on recommender systems. ACM, pp 165–172

Morrison PD, Roberts JH, Von Hippel E (2000) Determinants of user innovation and innovation sharing in a local market. Manage Sci 46(12):1513–1527

Neuhofer B, Buhalis D, Ladkin A (2014) A typology of technology-enhanced tourism experiences. Int J Tourism Res 16(4):340–350

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. In: Human language technologies: the 2010 annual conference of the North American chapter of the association for computational linguistics, HLT '10, pp 100–108. Stroudsburg, PA, USA

Oliveira P, von Hippel E (2011) Users as service innovators: the case of banking services. Res Policy 40(6):806–818

Olson EL, Bakke G (2001) Implementing the lead user method in a high technology firm: a longitudinal study of intentions versus actions. J Prod Innov Manag 18(6):388–395

Parasuraman A, Zeithaml VA, Berry LL (1988) Servqual—a multiple-item scale for measuring consumer perceptions of service quality. J Retail 64(1):12–40

Rihova I, Buhalis D, Moital M, Gouthro M-B (2015) Conceptualising customer-to-customer value co-creation in tourism. Int J Tourism Res 17(4):356–363

Rokach Lior, Maimon Oded Z (2014) Data mining with decision trees: theory and applications, 2nd edn. World Scientific Publishing Co., Inc., River Edge, NJ, USA

Rossetti M, Stella F, Zanker M (2016) Analyzing user reviews in tourism with topic models. Inf Technol Tourism 16(1):5–21

Sawhney M, Verona G, Prandelli E (2005) Collaborating to create: the internet as a platform for customer engagement in product innovation. J Interact Mark 19(4):4–17

Schwarz E, Faullant R, Krajger I, Breitenecker R (2009) Are leading edge users creative. In: Annual meeting of the 38th EMAC conference

Sigala M (2010) Managing customer involvement and roles in new service development (nsd): lessons learnt from www.mystarbucksidea.com. In: 28th Eurochrie conference, Amsterdam, The Netherlands

Skiba F, Herstatt C (2009) Users as sources for radical service innovations: opportunities from collaboration with service lead users. Int J Serv Technol Manage 12(3):317–337

Sternberg RJ, Lubart TI (1999) The concept of creativity: prospects and paradigms. Handb. Creativity 1:3–15

Urban GL, Von Hippel E (1988) Lead user analyses for the development of new industrial products. Manage Sci 34(5):569–582

Verona G, Prandelli E, Sawhney M (2006) Innovation and virtual environments: towards virtual knowledge brokers. Organ Stud 27(6):765–788

Von Hippel, E. (1978). Successful industrial products from customer ideas. J Mark, 39–49

Von Hippel E (1986) Lead users: a source of novel product concepts. Manage Sci 32(7):791–805

Von Hippel E (2005) Democratizing innovation: the evolving phenomenon of user innovation. Journal für Betriebswirtschaft 55(1):63–78

Waldhör K, Rind A (2008) Etbloganalysis: mining virtual communities using statistical and linguistic methods for quality control in tourism. Inf Commun Technol Tourism 2008:453–462

Wallach HM, Murray I, Salakhutdinov R, Mimno D (2009) Evaluation methods for topic models. In: Proceedings of the 26th annual international conference on Machine learning. ACM, pp 1105–1112

Wang Y, Yu Q, Fesenmaier DR (2002) Defining the virtual tourist community: implications for tourism marketing. Tour Manag 23(4):407–417

Wang H, Lu Y, Zhai C (2010) Latent aspect rating analysis on review text data: a rating regression approach. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 783–792

Wellner K (2015) User innovators in the silver market: an empirical study among camping tourists. Springer

Yoo KH, Gretzel U (2008) What motivates consumers to write online travel reviews? Inf Technol Tourism 10(4):283–295

Chapter 9
Customer Data and Crisis Monitoring in Flanders and Brussels

Steven Valcke

Abstract Visit Flanders promotes Flanders (the northern part of Belgium) and Brussels internationally. Visit Flanders has experimented with big data in different forms: web monitoring (all publicly available sources, including social media), flight data, mobile data, scraping hotel review scores, and credit card data. We have learnt about the possibilities and challenges with interpreting and using them effectively. This article focuses on crisis monitoring and how Visit Flanders has used web monitoring data and flight data for market segments. Brussels was linked to terrorism in November 2015 when the Paris attacks happened and in March 2016 when Brussels was facing a terrorist attack. Fast reactions were required, based on fast data. In 2015 only web monitoring data were available at Visit Flanders. This query was reused in 2016 and flight data was also monitored. More intensive monitoring was needed then because of the greater March 2016 tourism threat. Big data helped Visit Flanders to react separately to individual markets.

Keywords Web monitoring · Crisis monitoring · Destination · Big data · Flight data · Market segmentation

9.1 Big Data at Visit Flanders

Visit Flanders works with several systems that process customer data, delivered by different providers.

- Web monitoring (all public available sources including social media) for social listening, brand monitoring and campaign monitoring
- Flight data to monitor campaigns and research booking behaviour
- Mobile data for measuring the success of individual events
- Scraping hotel review scores to monitor the visitor satisfaction

S. Valcke (✉)
Marketing Office, Grasmarkt 61, 1000 Brussels, Belgium
e-mail: steven.valcke@toerismevlaanderen.be

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_9

- Credit card data for spending
- Facebook data for profiling.

The web monitoring and flight data are discussed here because they were most important for crisis monitoring.

9.2 Web Monitoring Tool at Visit Flanders

The web monitoring system Synthesio used by Visit Flanders gathers online social media. It offers global coverage and querying in 80+ languages. It can be queried with Boolean operators. Synthesio analyses topics, types of contributor, media types, time, reach, and sentiment. The sentiment component is determined algorithmically. A social reputation score is calculated from sentiment, volume and influence, with the maximum value being 100. The system also has some practical and technical limitations, such as the following.

- It is important to understand the Boolean query syntax to design effective queries.
- Private social media conversations are not included, such as those on private Facebook pages. Public Facebook pages tend not to be owned by typical people.
- Not all information is gathered from the websites included. For example, not all TripAdvisor information is included because every attraction has its own page. People don't often mention the attraction or destination because they are on that page, so it is not in the text and will not be found. Review scores are not extracted by the system.
- The number of mentions is partly country specific. In some countries there is more interest not only because of the number of inhabitants but also because of the use of social media and the number of media. This is important when countries are compared.

9.3 Use of Flight Data at Visit Flanders

Visit Flanders uses flight data delivered by Amadeus. The data are origin—destination based, with the destination Brussels. There are daily data about booking and boarding and from departure and return flights stay lengths are derived. The data can be broken down by channel type (e.g., traditional travel agent, online travel agent) and purpose. Four benchmark destinations are also included (Paris, Amsterdam, Vienna and Copenhagen). System limitations include the following.

- Direct bookings are not included, such as for most low cost airlines.
- It is only about flights. The main markets for Flanders are the neighbouring countries: The Netherlands, France, UK and Germany. Most visitors from those markets arrive by car or train.

- Not all travelers who use a plane and visit Flanders arrive at Brussels airport. Long haul markets can use another airport and make a roundtrip passing through Flanders. For one market, one of the benchmarks is a competitor while on another market (long haul) it is instead an entry for Flanders. This is the case for Amsterdam and Paris. For many long haul trips, Flanders is combined with France or the Netherlands and often Paris and Amsterdam are used as entry airports.
- There is little valid flight data for some markets, such as China, which has its own booking system.

The data give Visit Flanders a better understanding of the booking behaviour in most markets where it is active. They show when and which channel is used for booking so planning and channel choice of conversion focused marketing actions can be finetuned to raise efficiency.

The data can also give an indication of the effect of this type of action. Web monitoring can be added as a supportive indicator because people talk about planning a holiday. There should be a certain sequence: buzz about planning a holiday, booking the flight, boarding on the flight and buzz when at the destination.

9.4 Crisis Situations

In November 2015 the Paris attacks happened and Brussels was linked to terrorism. In March 2016 Brussels was hit by terrorist attacks at the airport and in the metro. Both places are very important for tourists. These threatened the reputation of Brussels as a safe destination. Monitoring was necessary to be able to minimize reputation damage in each market. For this quick, market-sensitive data were needed. In 2016 we developed an approach that took market specificity into account. We divided our markets into three activity levels:

1. Nurture: Only pro-active B2B communication focusing on clarifying the situation. No marketing.
2. Communicate: Positive B2B and B2C communication and stories about the destination. No call to action or sales promotions.
3. Activate: All normal activities are restored and some extra actions focused on short term conversion are taken.

The markets were evaluated weekly during the first two months. By then, most markets were at the activate level and the evaluation frequency was reduced. The evaluation used the following.

- Web monitoring for reputation impact: Knowing if the situation is still talked about and how strong the conversation is.
- Flights booking data: Bookings are used as an indicator of the impact of reputation on conversion.
- Local office expert evaluations of the situation, reputation and conversion.

The overview of the level of each market was constantly available online for the local Flemish sector.

9.5 Web Monitoring for Crisis Management

Crisis web monitoring was set up in November 2015. A query was designed to track feelings of insecurity towards Flanders rather than about the Paris attack itself.

9.6 Query

A query was designed to monitor enough social media to get insights into what the target markets were thinking about Flanders. The query, called the Brussels safety query, combined the main Flanders destinations with some generic words expressing feelings of insecurity. With this general query, it was possible to start tracking immediately and compare the situation with November to get a sense of the impact.

> ("flanders" OR "flemish" OR "belgium" OR "belgian" OR "mechelen" OR "antwerp" OR "bruges" OR "brussels" OR "leuven" OR "ghent") AND ("unsafe" OR "safety" OR "cautious" OR "anxious" OR "dangerous)

Only the Boolean operators AND and OR were used. Because the NOT and NEAR operators were not used, the data included many irrelevant posts that a refined version may have eliminated. The query was translated into 13 languages: English, French, German, Dutch, Spanish, Italian, Swedish, Norwegian, Danish, Portuguese, Russian, Chinese and Japanese. These languages cover most Flanders markets to generated market-specific insights.

9.7 Processing and Interpretation

9.7.1 Volume

The number of mentions was the main indicator. It revealed when the insecurity conversation was over so that information about the situation was no longer a priority for a specific market. The number of query matches and standard deviation were determined based on the week before the event happened, calculating the average number of data per day and excluding outliers. That has been a normal week with low and mostly stable numbers. After the crisis it appeared that the estimated levels were correct since they returned to that level.

It was considered that the buzz had stopped when the number of mentions was on a normal level for three consecutive days. The normal level was taken as any volume within a standard deviation of the prior average.

9.7.2 Sentiment

Market sentiment suggests the size of the insecurity challenge. The sentiment is delivered through a social reputation score (SRS) which is calculated by the provider taking sentiment of the mention and influence of the source into account.

9.8 Results

9.8.1 Total Number of Mentions

Figure 9.1 shows the evolution of the number of mentions matching the query for a period of 19 days, starting the day before the event (both March 2016 and November 2015). In March 2016 the number of mentions reached a peak on the day of the event and dropped after that first day to a level that was higher than before. Although the level on day 3 was much lower than on day 1 it was still 9 times higher than on day −1. It took until day 12 before it started to drop to a normal level, which took a further 2 days. The smaller peaks in between are because of events related to the terrorist attack.

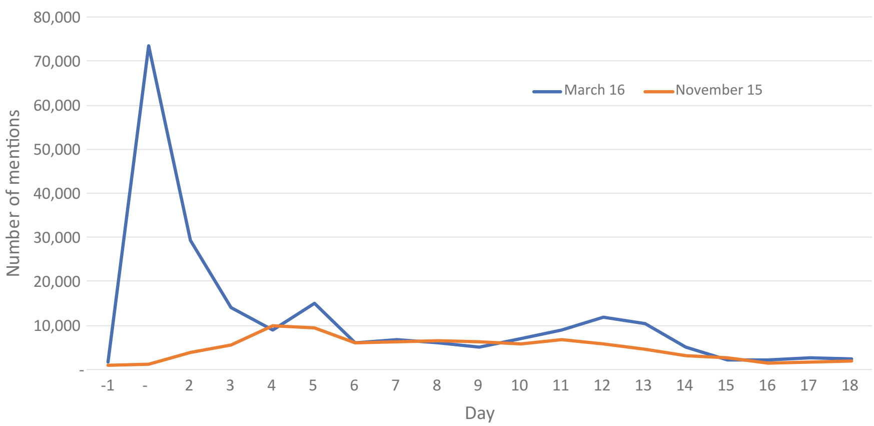


Fig. 9.1 Evolution of number of mentions matching Brussels safety query, day 0 is the day of attack. *Source* Synthesio

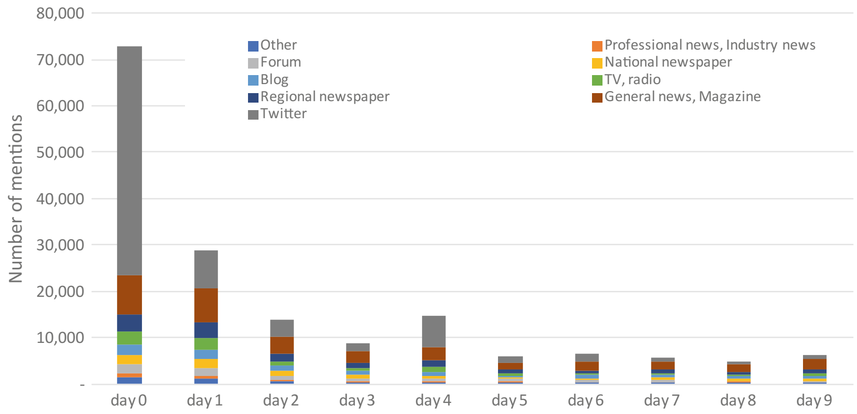


Fig. 9.2 Number of mentions matching Brussels safety query after the attack of March 22nd 2016 by media channel. *Source* Synthesio

The November 2015 pattern has some similarities. Brussels was in the news late on day 2 and for most people the focus was on day 3, causing a peak. From that point there was a similar evolution to March 2016.

The peak on day 0 was mainly driven by Twitter. People engaged emotionally with the content. From day 3, only traditional media were still reporting and the event was still in the news. The intermediate level is the level in which the event is still in the news but people don't engage any more (Fig. 9.2).

9.8.2 National Market Segments

Some countries are discussed here to illustrate the results. The day the normal level was reached depended on the country. In France this happened on day 14. The sentiment in France was also extremely negative. Although the story left the news relatively quickly, the destination brand still suffered. The same level of negativity was seen in Switzerland, a partly francophone country, presumably influenced by French media.

In Spain the normal level was only reached on day 31. Compared to 2015, the level seems relatively low because the events in Brussels in November 2015 were important news. In that week a football game between Spain and Belgium was planned in Brussels and cancelled. Sports always bring emotion and drive buzz. Events like football matches can enforce the impact.

In Japan it took even several months to return to normal level. Some caution when restarting business was required.

Russia and Sweden were atypical. In both countries the attacks were quickly out of the news but in both cases something happened in relation to the attacks to return them to the news. In the case of Sweden, one of the terrorists was found to have a link with the Swedish city of Malmö. In Russia, on day 10 it returned into the news because Putin mentioned it in one of his speeches and from then on it took 11 days to disappear. For social reputation, Sweden had a relatively low score, while in Russia it was more factual, with an average score of around 50 (Tables 9.1 and 9.2).

9.9 Flight Data in Case of Crisis Management

The flight data were an important element in the evaluation of the situation in each market. The data revealed how hard the destination was hit in terms of number of tourists and it told us when a market started to recover.

9.10 Processing and Interpretation

The flight data were analysed weekly. There were too many fluctuations from day to day to analyse it daily. Only the data concerning leisure travel were used. Leisure and business markets are different and could react differently to crises. The two segments within business, group and individual business travel could react very differently. It is not possible to distinguish the two segments in the data, which made it difficult to work with the business figures although Visit Flanders works on the MICE segment as well.

For effective interpretation it is important to be aware of how the number of bookings fluctuate naturally throughout the year because a decrease can be inherent to the period of the year and have nothing to do with the terrorist attacks. This differs by market. Easter plays an important role in some markets. Around Easter people travel and don't book but it is often also the end of the main booking period, but the date of Easter changes.

Another important fact to be aware of is that people cancelled their trip which caused retroactively a decrease in number of bookings in the weeks before the week of the 22nd March. The week of the 22nd March is in most markets the end of the high booking season so a decrease compared to this week can be expected.

To deal with the situation and the natural fluctuations, the ratios were converted into a score from 1 (best) to 5 (worst) and the average of both scores was used.

9.11 Year Over Year Approach

The first approach was based on the ratio between the number of bookings to Brussels in a certain week in 2016 compared to the number in the same week the year before for every relevant market. This approach eliminated the problem of the fact that

Table 9.1 Number of mentions matching safety query per day, starting from 22nd March 2016

	France	Spain	Japan	Russia	Sweden
Average without outliers	420	163	16	54	6
St dev without outliers	110	46	4	18	2
Day 0	10,979	2638	593	238	380
Day 1	3554	2062	489	164	126
Day 2	1655	1119	183	83	59
Day 3	1239	753	145	86	44
Day 4	2549	987	118	100	37
Day 5	747	856	55	46	42
Day 6	650	641	112	56	10
Day 7	797	588	80	70	36
Day 8	534	509	128	75	15
Day 9	831	745	136	59	28
Day 10	1256	1344	125	101	28
Day 11	1565	1263	103	114	8
Day 12	1487	1130	49	101	3
Day 13	694	732	37	125	4
Day 14	241	193	40	108	0
Day 15	260	248	36	102	6
Day 16	362	310	41	219	2
Day 17	318	166	30	102	16
Day 18	325	182	31	109	15
Day 19	161	189	28	87	32
Day 20	230	176	34	73	13
Day 21	264	226	33	28	5
Day 22	395	226	28	51	6
Day 23	339	204	32	28	5
Day 24	496	390	24	31	2
Day 25	126	194	21	10	2
Day 26	171	107	23	18	20
Day 27	176	141	39	29	5
Day 28	166	210	14	26	4
Day 29	233	220	33	43	8
Day 30	180	221	41	37	1
Day 31	329	127	32	20	3

Source Synthesio

Table 9.2 Social reputation score and comparisons for period of March

Country	Srs safety (/100)	Day 0 versus Day 8 factor	Mar 16 versus Nov 15 factor
France	21	21.71	2.76
Spain	57	6.35	2.01
Japan	na	10.37	2.09
Russia	46	3.26	1.91
Sweden	33	25.07	1.96

Source Synthesio

Table 9.3 Scoring system for the one year approach

Ratio n week x_y/n week x_{y-1}	Score
>1	1
1–0.9	2
0.9–0.7	3
0.7–0.4	4
<0.4	5

the main booking period was over and it was natural that there was a decrease. The aforementioned "Easter problem" might have caused some troubles but this approach still gives a good indication. The scoring used the Table 9.3 scale.

9.12 "Within Year" Approach

A second approach was needed. This approach worked only with weekly data from one year, 2016. For this approach a reference week is needed. Comparing with the week of the terrorist attack was logical but difficult because the evolution compared to this week depends on how hard the market was hit. A hard hit would be in favour for a quick recovery while markets almost not hit would have been evaluated as recovering slowly, certainly taking into account that in a lot of markets there is a natural decrease in the period after Easter.

The reference week was week 3 of the year. Week 3 is in all markets part of the main booking period for a coming year and was the least affected because the trips linked to these bookings were partly made already. This was the week when bookings were the best indication for what would have been a normal year. The closer to the attack the more cancellations impacted the number of bookings in the system. The scoring used the Table 9.4 scale.

Table 9.4 Scoring system for the within year approach

Ratio n week x/n week 3	Score
>1	1
1–0.9	2
0.9–0.7	3
0.7–0.4	4
<0.4	5

9.13 Results

As background information, the airport was closed for a couple of weeks, but tickets to Brussels could be booked during the whole period. The results here are for the three markets considered above: Spain, Russia and Sweden. France is not taken into consideration because flights from France to Brussels are not relevant. Similarly, Japanese tourists visiting Flanders and Brussels tend to use an airport in a neighbouring country.

The three markets show all a strong recovery in both indicators in the week after the attacks, but from then on differed. The boost the week after seems to be logic because on the first days after the attacks people won't think about going to Brussels so it is very low. A first small quick recovery is also to be expected because there are always people who really need to fly for one or another reason.

The Spanish market recovered very slowly. Twelve weeks after the attacks the number of bookings were still 40% lower compared to the same moment the year before and showed almost no increase. The slow recovery is consistent with the buzz analysis which showed that it took a long time until the buzz about the attacks had stopped. After 12 weeks there was no real sign of any improvement.

The Russian market was a bit more complex to interpret because it was also a market in an economic crisis. The first two weeks after the attacks (week 14 and 15) showed a strong increase. In week 16 there was a drop. This was at the same moment of the peak in online buzz. There seems to be a relationship between booking and online buzz. A second fact is that using the evolution within the year (fixing week 3) could be misleading. The "fix week 3" graph shows a recovery from week 20 on but this is not the case compared to the same period the year before. The increase is caused by a booking period for the May holidays in Russia. An increase is normal but the comparison to the year before shows that there wasn't a full recovery.

The Swedish market recovered. There was an increasing trend for both indicators. The drop at week 16 is in the week that it became clear that one of the terrorists had a link with Sweden and the online buzz started again. After week 16 the recovery restarted until week 21 but two weeks later the recovery was very strong and in week 25 both indicators were on or above the reference level again. The Swedish market could be considered as completely recovered after 12 weeks (Figs. 9.3, 9.4 and 9.5).

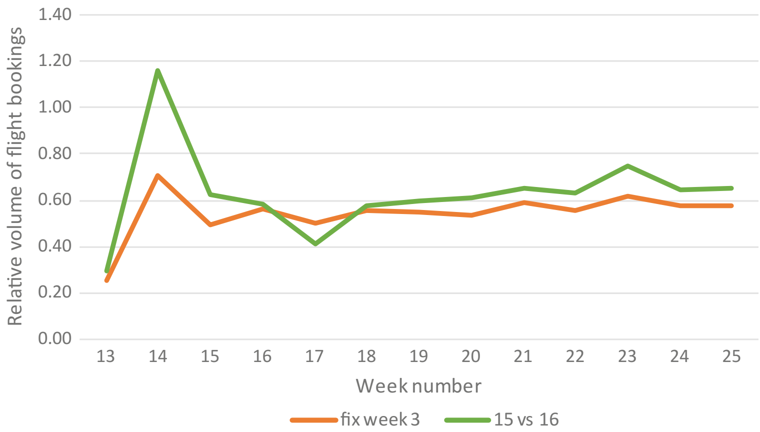


Fig. 9.3 Ratios based on flight data; Spanish market towards Brussels 2016. *Source* Synthesio

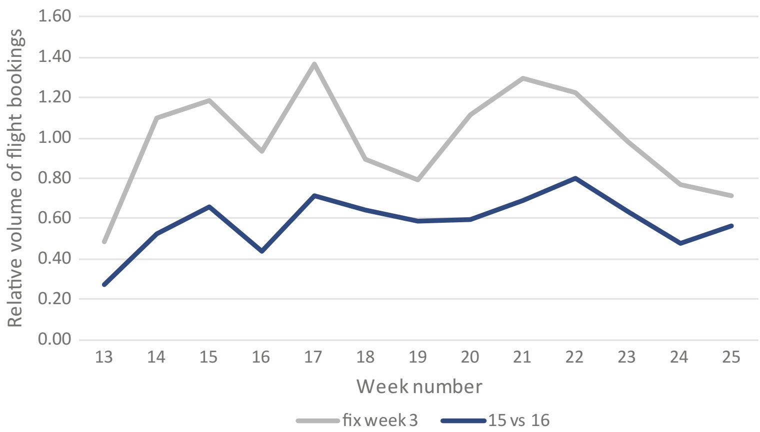


Fig. 9.4 Ratios based on flight data; Russian market towards Brussels 2016. *Source* Synthesio

9.14 Conclusion

Working with social media data, experience and knowledge of the potentials and limitations of the data is required to devise effective strategies. It is important to create reference points and an analysis strategy to generate actionable information. The strategy must ensure that the decisions that need to be made are backed by targeted data.

The example in this paper used social media data and flight bookings to decide on how to react to Flanders international tourism markets. The core web monitoring strategy focused on the volume of discussion about the selected key issue (insecurity in relation to Flanders). This appeared to be enough to make decisions about the level

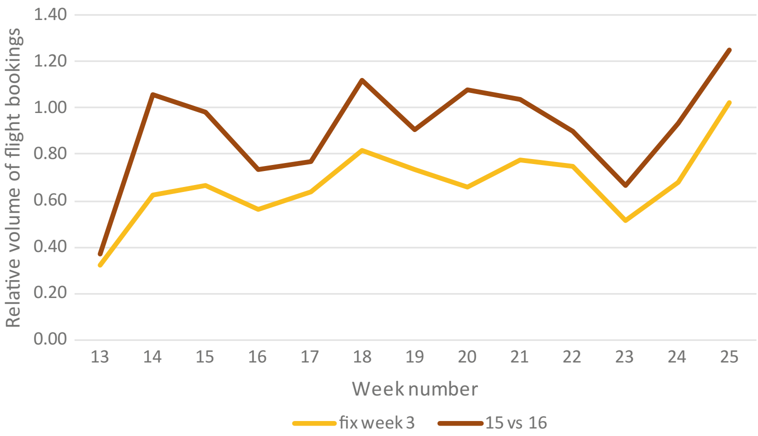


Fig. 9.5 Ratios based on flight data; Swedish market towards Brussels 2016. *Source* Synthesio

of concern in the target markets. The flight data were analysed against two reference points, the number of bookings in week 3 of 2016 and the number of bookings of the same week the year before. Both analyses have advantages and disadvantages. It was appropriate to combine both.

The examples confirm that markets react differently to events. While the Swedish market recovered very fast, the Spanish market was very slow. So it is logical that the communication strategy should be market-specific.

Chapter 10
Analyzing Airbnb Customer Experience Feedback Using Text Mining

George Joseph and Vinu Varghese

Abstract The objective of this chapter is to present a case of text mining on Airbnb user reviews to analyze and understand various aspects that drive customer satisfaction. The study can be extended further to discover segmentation and targeting of spaces that can take customer satisfaction to the next level and can also consider possibilities of geography specific, travel and purpose specific guest and host requirements. We are trying to gain insights about the challenges faced by customers in sharing economy along with ways to develop "super hosts". Thus, this work will try to advance our understanding about tourism and hospitality industry by presenting a case of big data analyses on Airbnb user reviews.

Keywords Text mining · Airbnb · Sentiment analysis · Rapid miner · Customer feedback

10.1 Introduction

The sharing economy has enabled people to do business in different ways. It has equipped the owners to monetize under-utilized assets over the internet through fee-based sharing; while it has also helped the guests use the services conveniently without compromising on quality and price. This has been proved by the ever-growing popularity of service providers like Airbnb, Uber, Lyft etc. The hospitality sector is now witnessing unparalleled growth with the advent of internet enabled technologies thereby getting rid of the traditional backwardness (Šerić et al. 2014). A plethora of opportunities and challenges presented to the guests, hosts and regulators make it a hot spot of interdisciplinary research for analytics, marketing, law, public policy and information technologies. Consequently, customer feedback which lies at the core of any industry is being increasingly studied for improving operations. It is in

G. Joseph (✉) · V. Varghese
DDUKK, CUSAT, Cochin, India
e-mail: george.joseph@roehampton-online.ac.uk

G. Joseph
University of Roehampton, London, UK

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_10

this context that tools like Rapid miner, which can analyze large volumes of texts are being increasingly employed to generate insights. Here in this chapter, we look closely at Airbnb, which is an online market place in the hospitality sector which facilitates listing and booking unique lodging spaces. This has led to the development of a community centered around public reviews of guests and hosts.

Founded in 2008, Airbnb acts as an online market place for hospitality service to rent short term lodging. As a broker between guests and hosts, the firm receives a percentage as service fee from both parties. Airbnb can be accessed through websites or mobile applications, wherein the guests and hosts can leave public reviews after the period of lodging. Guests can rate features of their stay including location, cleanliness and personal communication and thus give away important information for future hosting and stay. These reviews are published through the company website and involves details regarding destination, travel dates, and party size with description about attributes, photographs, payments and reviews from guests.

The present study aims to analyze guests' feedback using text mining with the following set of objectives:

- To understand the positive and negative aspects reflected in Airbnb reviews.
- To study the relative significance of positive and negative aspects revealed.
- To study the correlation between aspects in the reviews.

10.2 Literature Review

10.2.1 Customer Surveys and Text Mining

As stated by Xiang et al. (2017), research in tourism and hospitality sector has traditionally focused on surveys and interviews of stakeholders. This industry is crucial for the economy and guest satisfaction has become one of the critical measures of hospitality effectiveness and performance (Xiang et al. 2015). Guest satisfaction evaluated across time through consumer surveys or focus group interviews often suffers from poor sampling and low response rates resulting in vague assessments. The volume of data created every second in tourism and hospitality industry recommends the application of Big Data, a revolutionary phenomenon (Hassani and Silva 2015) involving the use of datasets that are so large and complex that traditional data processing applications and software tools are inadequate to capture and process within a reasonable period (Snijders et al. 2012). The purpose of this chapter is to show how to do text mining and sentiment analysis of customer data available online using RapidMiner software.

10.2.2 Online Reviews

Travelers using Airbnb generally conduct online search about destinations, hosts based and study reviews to make their travel and stay decisions. According to two surveys involving more than 2000 American adults (Pang and Lee 2008), more than 70% reported that online reviews had significant influence on their purchase and among them more than 32% had provided ratings via online ratings systems and about 30% of adults had posted online comments or reviews. Furthermore, it is stated that consumers are willing to pay more for an extra star-rated item. This proves the impact of online reviews on the image of destinations and hosts (Choi et al. 2007; Govers and Go 2005; Pudliner 2007). In addition, Choi et al. (2007) argue that the feedback available in travel blogs are richer in content and detailed, making them more popular than Likert-scale based survey results. These prove the effectiveness of word-of-mouth communications as a marketing strategy. Goldenberg et al. (2001) argues that customer's decision-making process is strongly influenced by word-of-mouth especially in the experience driven travel industry (Litvin et al. 2008).

10.2.3 Sentiment Analysis

Opinion mining or sentiment analysis is the mining of opinions, appraisals of individuals and their feelings towards objects, facts and attributes. In the case of Airbnb, there is a growing market focus on sustainability, law and customer issues which require industry and academia to increase their focus on individual host services. Reviews' positive and negative sentiment and the overall sentiment, consistency etc. are important factors on which Airbnb hosts can be appropriately prepared to take operational and strategic improvements. We will dig into the results to derive hosting insights for improving Airbnb operations strategy. Text mining as a methodology makes use of computational algorithms to identify patterns in huge textual data sets. The next part will provide information regarding analytical methods by describing sample, statistics and findings.

10.2.4 Airbnb Data

Airbnb generates millions of user review data for the public to analyze and gain new insights, transforming the hospitality industry like never before. In this case we used downloaded CSV files from the AirDNA website that contained millions of reviews stored in numerous files. This data set can provide insights if we are able to identify relationships among words in sentences of reviews (Upshall 2014). The data can include text, video and images. The semi-structured data can therefore be characterized to have a structure that is identifiable and analyzable using automated algorithms.

10.2.5 Tool

RapidMiner can do text mining using automated techniques in which computational algorithms extract meanings and patterns from texts by data preparation, machine learning and predictive model deployment. RapidMiner has many ready to use operators which can carry out text mining processes in stages. This requires the user to define the concepts and the context followed by importing the data, allocating dictionary, analysis and finally visualization (Gémar and Jiménez-Quintero 2015).

10.2.6 Case Study Procedure

To illustrate how text mining can be used, Airbnb data from London, available in AirDNA website as .CSV files was used. We employed 'AYLIEN' the Aspect Based Sentiment operator after setting the input attribute to "Review" and opting the domain "hotel" in the parameters. The parameter "hotel" resembles the Airbnb environment to a large extent in terms of measured attributes. We considered 1000 customer reviews (limited by the free AYLIEN version) made between April 9, 2014 and February 26, 2017. We were interested in analyzing the reviews that (1) were written in English, (2) of hosts based in London and got reviewed during the specified period and (3) were not duplicates. The version of the tool used is RapidMiner Educational version 8.00. Here we analyzed customer opinions using the free version of AYLIEN extension capable of Aspect-Based Sentiment Analysis as available in RapidMiner marketplace with a free account. The ABSA enables aspect and corresponding sentiment extraction from texts. There are trained domain specific models for industries such as restaurants, airlines, automobiles etc. The analysis process begins by identifying the top aspects and their sentiment followed by correlation analysis on the words, aspects and final visualization of findings.

10.2.7 Analysis

A new ExampleSet was generated under the Results tab and it displayed aspects present in reviews with their polarity. Before tokenization, we duplicated data using a Multiply operator to run two analyses in parallel. Word vectors were developed by tokenization with Process documents operator. For the first process, we tokenized

ABSA results in the format "aspect: polarity" and assigned weights to the newly created columns based on Binary Term Occurrences.

For the second processor, we used the duplicated set by tokenization to create unigram tokens, then transformed them to lowercase and removed tokens containing non-letters using a regular expression ([A-Za-z]*). Finally, we discarded tokens shorter than 3 characters and removed all stop words. Split operator was then used to separate out aspects and polarity. In the parameters section, the attribute to split and the split pattern was chosen. The Example Set generated attribute split into Aspect and Polarity columns. Filtered Example Sets show the positive and negative aspects and their counts. We used a Sort operator to get results in descending order of occurrence. In the final step we ran a correlation analysis to see common words that express particular sentiments towards aspects. For this, we first joined the two Example Sets forming one dataset with the words and the aspect: polarity pairs. We assigned numerical IDs using Generate ID operator and then used the Join operator to merge Example Sets and fed the result to the Correlation Matrix operator to filter and identify words extracted from reviews that correlated with a certain aspect: polarity attribute and vice versa. Finally, visualizations of results were done using charts capabilities.

10.2.8 Findings & Discussions

As already mentioned, we used the downloadable dataset instead of web crawling extension of Rapidminer. Airbnb CSV file with 1000 reviews had no ratings attached and therefore the analysis involved pure text mining. Despite the small scale of the study limited to the basic text mining methods, several relevant information could be extracted.

The Fig. 10.2 (see Appendix) shows the positive aspects as revealed in the reviews with location receiving maximum positive reviews followed by room amenities, comfort, cleanliness, staff, food and drinks, quietness, beds, value for money, customer support, design, facilities, view, WIFI and payment. These fifteen items therefore received varied significance in reviews and therefore must receive varying priorities for improvement. Location is a Real estate related or hard factor and therefore can be argued as a difficult aspect to improve but stood as the most important aspect in customers' mind capable of generating positive reviews. Room amenities is yet again a hard aspect but can be improved based on the requirements received in the reviews and feedbacks. Comfort requires further study to unearth the factors considered and the choices made. Cleanliness remains important in the shared facilities and will remain an operational challenge for hosts.

The Fig. 10.3 (see Appendix) presents the negative aspects with room amenities receiving the maximum negative reviews followed by cleanliness, location, beds, facilities etc. The change in importance of aspects is evident with location moving to fifth position. This can be due to the potential of room amenities, cleanliness to generate stronger negative sentiments compared to location. Another important finding is that the reviews have been predominantly positive with negative aspects receiving fewer mentions. For example, Room amenities received 221 positive mentions but garnered only 39 negative mentions.

The Fig. 10.7 (see Appendix) shows the correlation matrix. Positive reviews received for comfort were correlated to positive reviews received for beds with a coefficient of 0.414. Positive reviews for cleanliness were correlated to positive reviews received for comfort with a coefficient of 0.341.

The findings help us draw parallels from hotel industry. For example, cleanliness of rooms, service quality, and employees' knowledge and service are considered important in determining the satisfaction of hotel guests. Front line service providers in hotels are expected to provide a standardized, structured and simplified delivery process to satisfy their guests fully (Vijayadurai 2008). The customers using Airbnb may be influenced by the previous hotel experiences too. As Choi and Chu (2001) concluded, the staff quality, room qualities and value are important in determining guest satisfaction.

Thus, we can say that there are some aspects that can decide the net polarity of reviews common to Airbnb and hotel industry and therefore must be considered for in depth analysis in order to improve the level of customer satisfaction. This analysis revealed some common aspects and their rankings in terms of occurrences in the reviews. Location could largely bring positive reviews as against room amenities which received negative reviews. This calls for further analysis into sub factors that can be improved to bring consistent positive reviews. The other aspects like customer engagement and support in terms of active participation in hosting can strengthen a host's competitive position. Furthermore, it can enhance the goodwill and therefore increase profitability. An aspect mentioned in preliminary reviews tends to receive further ones asserting or negating the opinions. This again proves that consumer feedback has the power to shape opinions and behavior of future customers (He et al. 2013).

Thus, the interactions on website such as Airbnb bring aspects of the Airbnb hosts' customer service into public scrutiny. Firstly, Airbnb hosts need to track online conversation from a location and host specific manner. Airbnb can use the findings of text mining of competitive listings and benchmark 'super hosts' for giving insights to underperforming hosts by benchmarking (He et al. 2013). Airbnb must enable an automatic way of giving feedback based on benchmarked hosts' service in similar locality and price range which can provide insights on points-of-parity (POP) and Points of difference (POD). POPs here refer to areas in which a host is performing strongly (Krajnović et al. 2013). For instance, if a host discovers that consumers are dissatisfied with the beds and have admired those of a competitor in the locality with similar price range, then he or she could rectify the deficiency. Furthermore, by identifying common trends in the discussion, Airbnb can advise hosts on ways to differentiate themselves and help them develop points-of-difference (POD) to develop a strength, making their hosting stand out. Text analysis could uncover major themes of the reviews within the period. It should be noted that in terms of frequency of occurrence, findings help hosts to map the key trends and emphasize the relevant points in marketing and strategic decision making. The dynamic nature of the hospitality industry calls for continuous improvement of products and services (Peters and Pikkemaat 2006). Text mining is becoming part and parcel of the new paradigm in tourism and hospitality domains and studies based on Airbnb data can therefore help common public employed in the hospitality field.

This study contributes to theory and practice in several ways. Firstly, it provides a case of scalable aspect-based sentiment analysis process applied to specific location-based Airbnb units. This can help developing marketing strategies and therefore showcase how Airbnb hosts can increase competitive advantage and create 'collective intelligence' from customer databases. Airbnb has already started taking initiatives to consider novel trends and generalized opinions to improve business. Another contribution lies in the proposed method applied to an Airbnb unit in a location which may expose how guest satisfaction is being perceived, thereby providing valuable knowledge for Airbnb hosts to understand the strengths and weaknesses of comparable units in a locality. From a practical point of view, this study also enforces the core sentiments expressed in mainstream Airbnb issues.

Nevertheless, the present study comprises few limitations. First, no predictive modeling was done and the sample represents only a single Airbnb location. Therefore, the specific Airbnb issues identified in customer reviews might reflect the perceptions of location-related aspects of Airbnb hosts and guests that frequent the location. Sentiment classification and guest satisfaction could be considerably different in another geographical context. Another limitation relies on nonuse of clustering algorithms. This research however makes a practical case for the development of a methodology that can be practiced by individual hosts or made available to them by Airbnb or third party as a service characterized by customer focus and location based competitive intelligence. As such, it is expected that this study sets an example for the development of business intelligence systems that can be used and applied by the public. Even though customer feedback is vital in Airbnb's efforts toward continuous improvement, a comprehensive report on the customer experience remains a challenge and this paper considers this as a future research opportunity. Consequently, Airbnb can kick off some location specific initiatives if needed to address travelers' concern enabling segmentation, targeting and positioning.

10.2.9 Theoretical Implications

This study attempts to enrich the understanding of the hospitality in the sharing economy by examining the outcomes benefits customers seeks from service providers. Existing literature in relationship marketing suggests that customers primarily seek benefits from service exchanges (Gwinner et al. 1998; Dagger and Sweeney 2006). There is a correlation between sentiment in online reviews and sales (Hu et al. 2014). Therefore, it is logical to suggest that textual reviews contribute to consumers purchase decisions and justify the study. According to Kwok and Xie (2016), the value of online reviews depends on the receptiveness of employees and resulting financial performance (Xie et al. 2017). Other main directions of theoretical implications are towards enhancing unique customer value by increasing benefits or reducing sacrifices (Zeithaml et al. 1998) and build high-value dimension that competitors find difficult to imitate (Rintamaki et al. 2007). Based on current findings, suggestion for further research could focus on operational mechanisms of review system as well as its effect on customer choices.

10.2.10 Practical Implications

We could demonstrate how text-mining methods can extract valuable information embedded in reviews. These texts can give insights enabling Airbnb hosts better meet customer expectations. This study hails the use of freely available data source which is location specific and hence recommends taking location specific studies in improving the competitiveness of hosts. Travel review applications are getting more and more popular with free online tools and increasing accessibility to text mining technology. Although this case study was based on Airbnb facilities located in London, it can be extended to Airbnb hosts everywhere as Rapid Miner software has features to process text in many languages with self-made dictionaries. Thus, the software can be used to process content in several different languages, opening opportunities internationally for businesses. The research seems to suggest that by screening Airbnb reviews and finding the most relevant discussion points within the data, Airbnb hosts can do a SWOT or competitor analysis. Furthermore, this is not limited to only finding out the trends in the feedback for a specific Airbnb host but also identifying themselves among the competitive listings. This means that hosts can map what competitors are doing better with the aim of replicating those aspects into their own service. A comprehensive study could be performed with a larger set of reviews in multiple locations along with the ratings. More popular and advanced text mining procedures like Latent Dirichlet Allocation in topic modeling could reveal more insights. The study also recommends developing word dictionaries customized for Airbnb data.

10.3 Conclusion

Airbnb has already impacted the hotel industry in various ways and therefore the latter is also looking forward on ways to accommodate Airbnb or incorporate the valuable customer feedback generated. In addition, the study demonstrates how basic text mining techniques can be used to conduct a content analysis with minimum technological expertise, hardware and software resources. Thus, hosts should familiarize themselves with such analysis to strengthen their position in the international business environment. As demonstrated, even basic text mining procedures can offer several insights and thus can benefit hosts in numerous ways. Automated text analysis technologies can help Airbnb hosts' 'position' in the international market place by continuous monitoring. Future researchers can help define the pathways that underlie the creation of extra ordinary, memorable experiences, which subsequently elicit favorable behavioral intentions. Therefore, text mining will continue to grow in popularity in hospitality industry.

Appendices

See Figs. 10.1, 10.2, 10.3, 10.4, 10.5, 10.6, 10.7, 10.8 and 10.9.

Row No.	in documents	total	word_1	word_2
1	12	12	beds	negative
2	2	2	beds	neutral
3	52	52	beds	positive
4	16	16	cleanliness	negative
5	4	4	cleanliness	neutral
6	134	134	cleanliness	positive
7	3	3	comfort	negative
8	3	3	comfort	neutral
9	160	160	comfort	positive
10	5	5	customer support	negative
11	4	4	customer support	neutral
12	39	39	customer support	positive
13	1	1	design	negative
14	37	37	design	positive
15	8	8	facilities	negative
16	4	4	facilities	neutral

Fig. 10.1 Aspects Polarity

Row No.	word	in documents	total
1	location:positive	298	298
2	room amenities:positive	221	221
3	comfort:positive	160	160
4	cleanliness:positive	134	134
5	staff:positive	102	102
6	food/drinks:positive	68	68
7	quietness:positive	62	62
8	beds:positive	52	52
9	value:positive	43	43
10	customer support:positive	39	39
11	design:positive	37	37
12	facilities:positive	34	34
13	view:positive	18	18
14	wifi:positive	10	10
15	payment:positive	1	1

Fig. 10.2 Positive Aspects Example set

Row No.	word	in docu...	total
1	room amenities:negative	39	39
2	cleanliness:negative	16	16
3	location:negative	16	16
4	beds:negative	12	12
5	facilities:negative	8	8
6	food/drinks:negative	7	7
7	value:negative	7	7
8	quietness:negative	6	6
9	staff:negative	6	6
10	customer support:negative	5	5
11	wifi:negative	4	4
12	comfort:negative	3	3
13	payment:negative	2	2
14	design:negative	1	1

Fig. 10.3 Negative Aspects Example Set

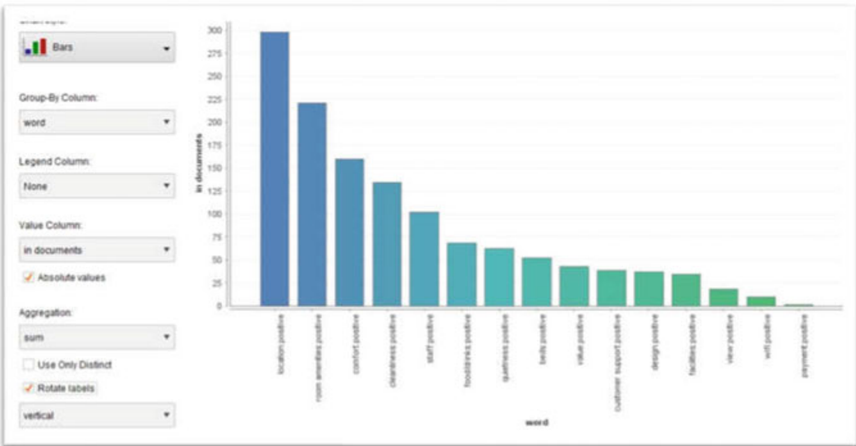


Fig. 10.4 Positive Aspects graph

Row No.	word	in docu...	total
1	room amenities:negative	39	39
2	cleanliness:negative	16	16
3	location:negative	16	16
4	beds:negative	12	12
5	facilities:negative	8	8
6	food/drinks:negative	7	7
7	value:negative	7	7
8	quietness:negative	6	6
9	staff:negative	6	6
10	customer support:negative	5	5
11	wifi:negative	4	4
12	comfort:negative	3	3
13	payment:negative	2	2
14	design:negative	1	1

Fig. 10.5 Negative Aspects graph

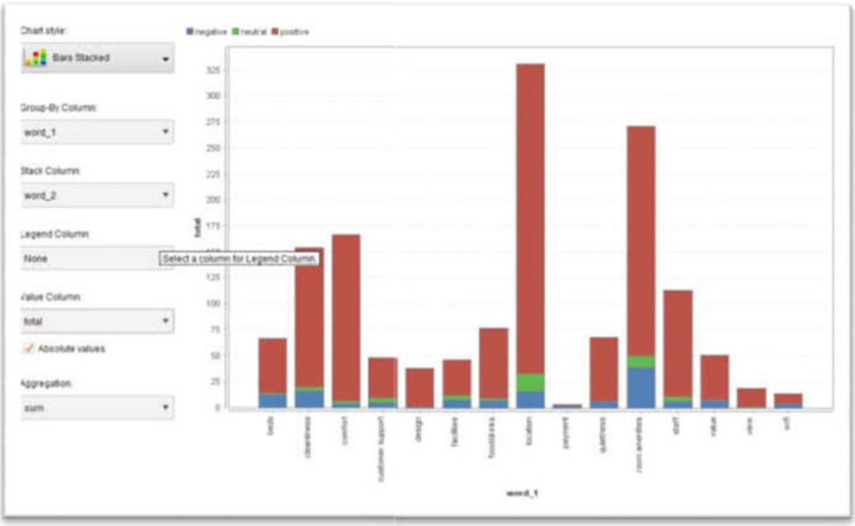


Fig. 10.6 Aspects Stacked

Attributes	Comme...	beds:ne...	beds:ne...	beds:po...	cleanlin...	cleanlin...	cleanlin...	comfort...	comfort...	comfort...	custom...	custom...
Comments	1	0.007	-0.012	-0.017	-0.011	0.020	-0.057	0.037	-0.010	-0.057	-0.012	-0.035
beds:negative	0.007	1	-0.005	-0.026	0.279	-0.007	-0.016	0.162	-0.006	0.052	-0.008	-0.007
beds:neutral	-0.012	-0.005	1	-0.010	-0.006	-0.003	-0.018	-0.002	0.407	-0.020	-0.003	-0.003
beds:positive	-0.017	-0.026	-0.010	1	-0.030	0.128	0.146	-0.013	-0.013	0.414	-0.017	0.057
cleanliness:negative	-0.011	0.279	-0.006	-0.030	1	-0.008	-0.050	-0.007	-0.007	-0.012	0.104	0.244
cleanliness:neutral	0.020	-0.007	-0.003	0.128	-0.008	1	-0.025	-0.003	-0.003	0.102	-0.004	0.247
cleanliness:positive	-0.057	-0.016	-0.018	0.146	-0.050	-0.025	1	0.032	-0.022	0.341	-0.028	0.022
comfort:negative	0.037	0.162	-0.002	-0.013	-0.007	-0.003	0.032	1	-0.003	-0.024	-0.004	-0.003
comfort:neutral	-0.010	-0.006	0.407	-0.013	-0.007	-0.003	-0.022	-0.003	1	-0.024	-0.004	-0.003
comfort:positive	-0.057	0.052	-0.020	0.414	-0.012	0.102	0.341	-0.024	-0.024	1	0.008	0.059
customer support:negative	-0.012	-0.008	-0.003	-0.017	0.104	-0.004	-0.028	-0.004	-0.004	0.008	1	-0.004
customer support:neutral	-0.035	-0.007	-0.003	0.057	0.244	0.247	0.022	-0.003	-0.003	0.059	-0.004	1
customer support:positive	-0.059	-0.022	-0.009	0.092	0.015	-0.013	0.088	-0.011	-0.011	0.039	-0.014	-0.013
design:negative	0.017	-0.003	-0.001	-0.007	-0.004	-0.002	-0.012	-0.002	-0.002	-0.014	-0.002	-0.002

Fig. 10.7 Correlation Matrix

Fig. 10.8 Positive Correlation Sample-Facilities

Attributes	facilities:positive ↓
facilities:positive	1
food/drinks:positive	0.519
location:positive	0.252
staff:positive	0.247
room amenities:positive	0.246
quiet	0.205
area	0.203
view:positive	0.182
restaurant	0.180
shop	0.157
london	0.139
quietness:positive	0.135
welcome	0.127
flat	0.121

Fig. 10.9 Negative Correlation Sample-Facilities

Attributes	facilities:negative ↓
facilities:negative	1
food/drinks:negative	0.531
view:neutral	0.352
staff:negative	0.284
room amenities:negative	0.272
location:negative	0.257
cleanliness:negative	0.167
apartment	0.154
kitchen	0.146
restaurant	0.130
walk	0.127
nice	0.107
bus	0.100
bathroom	0.096

References

Choi S, Lehto XY, Morrison AM (2007) Destination image representation on the web: content analysis of Macau travel related websites. Tourism Manage 28(1):118–129. https://doi.org/10.1016/j.tourman.2006.03.002

Choi TY, Chu R (2001) Determinants of hotel guests' satisfaction and repeat patronage in the Hong Kong hotel industry. Int J Hospitality Manage 20(3):277–297. https://doi.org/10.1016/s0278-4319(01)00006-8

Dagger TS, Sweeney JC (2006) The effect of service evaluations on behavioural intentions and quality of life. J Ser Res 9(1):3–18

Gémar G, Jiménez-Quintero JA (2015) Text mining social media for competitive analysis. Tourism Manage Stud 11(1):84–90

Goldenberg J, Libai B, Muller E (2001) Talk of the network: a complex systems look at the underlying process of word-of-mouth. Marketing Lett 12(3):211–223

Govers R, Go FM (2005) Projected destination image online: website content analysis of pictures and text. Information Technol Tourism 7:73–89

Gwinner KP, Gremler DD, Bitner MJ (1998) Relational benefits in services industries: the customer's perspective. J Acad Marketing Sci 26(2):101–114. https://doi.org/10.1177/0092070398262002

Hassani H, Silva ES (2015) Forecasting with big data: a review. Ann Data Sci 2(1):5–19

He W, Zha S, Li L (2013) Social media competitive analysis and text mining: A case study in the pizza industry. Int J Information Manage 33(3):464–472

Hu N, Koh NS, Reddy SK (2014) Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. Decis Support Syst 57:42–53. https://doi.org/10.1016/j.dss.2013.07.009

Krajnović A, Bosna J, Jašić D (2013) Umbrella branding in tourism—model regions of Istria and Dalmatia. Tourism Hospitality Manage 19(2):201–215

Kwok L, Xie KL (2016) Factors contributing to the helpfulness of online hotel reviews. Int J Contemporary Hospitality Manage 28(10):2156–2177. https://doi.org/10.1108/ijchm-03-2015-0107

Litvin SW, Goldsmith RE, Pan B (2008) Electronic word-of-mouth in hospitality and tourism management. Tourism Manage 29(3):458–468. https://doi.org/10.1016/j.tourman.2007.05.011

Pang B, Lee L (2008) Opinion mining and sentiment analysis. Foundations Trends Information Retrieval 2(1–2):1–135. https://doi.org/10.1561/1500000011

Peters M, Pikkemaat B (2006) Innovation in tourism. J Q Assurance in Hospitality Tourism 6(3–4):1–6. https://doi.org/10.1300/j162v06n03_01

Pudliner BA (2007) Alternative literature and tourist experience: travel and tourist weblogs. J Tourism Cultural Change 5(1):46–59. https://doi.org/10.2167/jtcc051.0

Rintamäki T, Kuusela H, Mitronen L (2007) Identifying competitive customer value propositions in retailing. Manag Ser Q 17(6):621–634

Šerić M, Gil-Saura I, Ruiz-Molina ME (2014) How can integrated marketing communications and advanced technology influence the creation of customer-based brand equity? Evidence from the hospitality industry. Int J Hospitality Manage 39:144–156. https://doi.org/10.1016/j.ijhm.2014.02.008

Snijders C, Matzat U, Reips UD (2012) Big Data: big gaps of knowledge in the field of Internet Science. Int J Internet Sci 7(1):1–5

Upshall M (2014) Text mining. Business Information Rev 31(2):91–99. https://doi.org/10.1177/0266382114541180

Vijayadurai J (2008) Service quality, customer satisfaction, and behavioral intention in hotel industry. J Marketing Commun 3(3):14–26

Xiang Z, Du Q, Ma Y, Fan W (2017) A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. Tourism Manage 58:51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Xiang Z, Schwartz Z, Gerdes JH, Uysal M (2015) What can big data and text analytics tell us about hotel guest experience and satisfaction? Int J Hospitality Manage 44:120–130. https://doi.org/10.1016/j.ijhm.2014.10.013

Xie KL, So KK, Wang W (2017) Joint effects of management responses and online reviews on hotel financial performance: a data-analytics approach. Int J Hospitality Manage 62:101–110. https://doi.org/10.1016/j.ijhm.2016.12.004

Zeithaml VA, Berry LL, Parasuraman A (1996) The behavioral consequences of service quality. J Marketing 60(2):31. https://doi.org/10.2307/1251929

Chapter 11
Big Data as a Game Changer: How Does It Shape Business Intelligence Within a Tourism and Hospitality Industry Context?

Nikolaos Stylos and Jeremy Zwiegelaar

Abstract With the advent of web analytics, data mining and predictive modeling, businesses have nowadays a better knowledge in creating more efficient and effective processes for meeting customers' needs, driven by a wealth of available information. The value of big data in influencing business intelligence in the tourism and hospitality industry has also been widely acknowledged, as the synergetic utilization of big data can enhance organizations' decision support systems to reach process optimization. Notwithstanding empirical research on exploring the implications of utilizing big data in the tourism sector has been published in the last few years, there is still need of a framework that would serve as the bedrock of taking the relevant conceptualization one step forward. Therefore, this chapter demonstrates the crucial role of big data in matching organizational objectives with tourist needs through delineating and detailing the analytical frameworks to support an advanced B2C interface, based on various internal databases and external data sources. The role of stakeholders and necessary resources are explained, and the full potential of big data in tourism and hospitality is revealed.

Keywords Big data · Analytical framework · Business intelligence · Marketing · Stakeholders · Tourism · Hospitality

11.1 Introduction

Data and databases are important resources for business units and organizations, including those of the tourism and hospitality sector (Fuchs et al. 2014; Marine-Roig and Clavé 2015). They are the building blocks for data-driven management, which is of utmost importance currently for informed decision making. This data may refer to tourism organizations' inputs and outputs, as well as key performance indicators

N. Stylos (✉) · J. Zwiegelaar
University of Bristol, Bristol, England, UK
e-mail: n.stylos@bristol.ac.uk

J. Zwiegelaar
e-mail: jeremy.zwiegelaar@bristol.ac.uk

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_11

(Buhalis and Foerste 2015; Gretzel 2015b). Additionally, consumer data regarding personal needs and preferences are collected regularly, as they are stored by users on digital platform profiles, which can greatly contribute in shaping tourism and hospitality product offerings that match individual characteristics. The combined use of data originating from various sources that would cover different aspects of individuals' lives is a reality now, through digital networks offering big opportunities for improved business competitiveness; this is the Big Data evolution.

The value of big data in shaping business intelligence in the tourism industry has been expanding lately with numerous researchers postulating that public and private tourism-related organizations have started moving from conventional database management to Big data (Buhalis and Foerste 2015; Gretzel et al. 2015a). The challenge for tourism organizations is to determine how to optimize the use of data available from various sources, stored in different forms and sometimes being inadvertently collected, to improve their decision-making processes (Gretzel et al. 2015a; Xiang et al. 2015). To respond effectively to this challenge and leverage any obtainable opportunities, it is important to unravel big data implementation and develop a systematic way for utilizing it. Big data should be developed to support tourism destinations and hospitality practice. That in turn would also influence the service offering to visitors and hotel guests, which may create new opportunities for evaluating customers' feedback on various aspects of the service delivery (Kunz et al. 2017; Liu et al. 2017; Miah et al. 2017). The premise of big data is to create a smart overarching synergetic framework, where individual data packets would integrate to an optimized system of tourism business intelligence to enhance organizations' decision support system. This is vital for service systems where value co-creation is the product of various resources, such as organizational policies, internal and external systems, proprietary technologies, human resources and, recently, artificial intelligence (Maglio and Spohrer 2008). In this vein, the main aim of this chapter is to illustrate how Big data formulate new ways for managers to improve tourism and hospitality organizations' capabilities. This would ultimately offer improved tourism and hospitality services and enhanced customer experiences, for the progress of the sector at large. Specifically, this chapter demonstrates the crucial role of big data in matching organizational objectives with customer needs by providing the analytical frameworks to support an advanced B2C interface by combining various internal databases and external data sources.

11.2 Business Intelligence in Tourism and Hospitality

Information and communications technology applications in tourism have been increased lately (Buhalis and Foerste 2015) and Big data is becoming a very promising vehicle to support decision making in tourism marketing and management.

Decisions within the organization are made from an awareness of the utility of data and the clear and deliberate strategy for organizing data within the organization i.e. having a clear strategy focused on data driven decisions. Improvement in organizational effectiveness requires transforming processes, changing corporate service ecosystems and enabling innovation (Brown et al. 2011; Manyika 2011). Service science and system innovation is the vehicle to conceptualize, categorize and understand the way service systems interact with each other to achieve value co-creation (Maglio and Spohrer 2008). Value co-creation offers the critical mass to link the successful implementation of a decision-making system with organizational performance. In line with this idea, the Service-Dominant (S-D) logic (Vargo and Lusch 2004) has been proposed as the conceptual foundation for building appropriate system theories for a better understanding of contemporary, dynamic service systems.

Various theories have been introduced to conceptualize service ecosystems, such as the complex network theory (Borgatti et al. 2009), social network theory (Goldenberg et al. 2001), and the multi-attribute value theory (Ferretti and Comino 2015) to improve decision-making via a dynamic configuration of organizational resources, including the available data flows. Bean and Kiron (2013) also proffers that the data-driven decision making is a promising trend and that the key factor for successful use of big data flows is the organizations' adoption of it in every aspect within their organization. Accessibility and availability of data is needed to ensure technological innovation is integrated to support organizational decisions. Within the tourism sector currently there are many organizations who are collecting, storing and analyzing data for strategic business decisions providing valuable knowledge. The ability to manage, analyze and act on data ("data-driven decision systems") is crucial to organizations and is seen as a significant asset. The prospects of big data analytics are important and the benefits for data-driven organizations are significant determinants for competitiveness and innovation performance. However, there are also considerable constraints for organizations to adopt data-driven approaches and gain the benefit of this valuable knowledge through big data.

Organizations of the tourism sector already engaged in using big data, report great **value** being captured within their structures and also between organizations (Phillips-Wren and Hoskisson 2015). The power of data is in what they are used to predict or show. The data themselves come to life and begin to have consequences when they are analyzed and when those analyses are integrated into social, governmental and organizational structures (Beer 2018). Tourism organizations use big data to analyze which variables provide the most interesting correlations and the significant ones are used to find and determine causality with models. Many firms of the sector, mainly large ones, have developed competitive advantage, leading to better **performance**, through the effective use of databases and available analytical capabilities (Gretzel et al. 2015b). Davenport and Harris (2007) found a positive correlation between higher levels of analytics used and 5-year compound annual growth rate for 32 organizations in their survey. Lavalle et al. (2010) found that organizations using business information and analytics for differentiation within their industry had twice the likelihood of being top performers as lower performers.

The abundant information that flows from various sources to different directions throughout the tourism sector stakeholders' system in the form of data packets encompasses various aspects of tourist activities. This individual-based tourist data apply and are supported by smart technology which equip the organizations with dynamic capabilities to ultimately be part of a smart tourism ecosystem (Gretzel et al. 2015a). A smart tourism ecosystem (STE) is therefore defined as being a tourism system that utilizes smart technology in developing, managing and the intelligent delivery of touristic services/experiences and is constituted by a desire for transparent sharing of information and value co-creation (Gretzel et al. 2015b). Guo et al. (2014) refer to this phenomenon of data transformation using big data as *informatization* of tourism. Because of smart technology integration, it offers a deeper insight into the evaluation of processes and feedback about tourist experience and organizations' value.

Furthermore, the penetration of big data in the tourism destination research has recently appeared in various publications including the work of Miah et al. (2017), who investigated the capability of big data to provide insights of tourist's behavioral patterns at a destination. The capability of the solution provided by big data was illustrated in a case study of inbound tourists to Melbourne in Australia (Miah et al. 2017). The investigators adopted design science research methods to design and evaluate a 'big data analytics' method to support strategic decision-making in tourism destination management. Tourists used geotagged photos uploaded to the social media site, Flickr. The ability of the method in showcasing destination management organizations to analyze and predict tourist behavioral patterns at specific destinations in Melbourne, Australia was shown as a representative case. The created artefacts provide a method for analyzing unstructured big data to provide deeper strategic decision making within a real problem domain.

11.3 Data Mining and Predictive Modeling

Big data is largely driven by Web analytics. This is particularly important for both business via sole online platforms and also physical ones that have an e-shop in parallel to the traditional channel. Web analytics reflect the process of taking customer views and traffic to facilitated reporting of sales and/or time conversion on digital platforms. The aim is to measure how the customers experience firms' web platforms and produce relevant reports. Thus, with web analytics reports, firms can improve the web platforms to meet consumers' expectations and, overall, better serve the customers shopping needs (Priporas et al. 2017). Often feedback is required from user testers on the look and feel of websites to ensure that links are working appropriately.

To make predictions based on the immense volumes of data available, tourism and multinational hospitality organizations have started conducting predictive modeling by employing a combination of algorithms and machine learning techniques (Morabito 2015; Phillips-Wren and Hoskisson 2015; Xiang et al. 2015). Tourism-related platforms, such as Tripadvisor, Booking.com and Trivago are already providing tourists with recommendations which cover various aspects of their travel experience. These platforms generate more than one-third of their sales by using sophisticated recommender systems to identify, rank and provide suitable product recommendations (Pantano et al. 2017; Xiang et al. 2015).

Big data consist of different data forms such as messages, images, geodemographics and updates retrieved from various digital sources and networks. Benefits of the analyses of these sources, e.g. social media tweets from Twitter, may provide enriched information about international tourists' intentions to visit a specific tourism destination, thus predicting relevant trends (Kunz et al. 2017; Zhang 2012). Consequently, Big data provide the opportunity to collect real-time unstructured and structured data of actual or potential tourists and hospitality guests pertaining to their behavior. That should enhance marketing communications, better segmentation, targeting and positioning of tourism product. Hence, big data are the medium to faster, more effective and meaningful tourism management and business performance to the benefit of stakeholders and the tourism economy at large (Morabito 2015).

11.4 Creating Better Knowledge by Using Efficient Processes

The data driven decision-making mindset, according to Davenport and Harris (2007), enables managers to make decisions based on a rationale from evidence taken from the big data. Decisions within the organization are made from awareness of the usefulness of data and the clear and deliberate strategy for organizing data within the organization (Phillips-Wren and Hoskisson 2015). The change requires transforming processes, changing corporate ecosystems and enabling innovation (Brown et al. 2011). Accessibility and availability of data is needed to ensure technological innovation is integrated to support organizational decisions. When there is a clear data driven strategy in place which becomes part of the organization culture, then there is better use of big data and thus more chances of adopting the integrated approach leading to organizational success.

Much of the current thinking about big data focuses on the usefulness and the potential that it offers for marketing business products and services. There are concerns about the use of big data within the software industry as to its usefulness when organizations have limited knowledge and skills or capability to exploit the opportunities posed. There are different understandings of the usefulness of big data as a result of varying definitions of the phenomenon. The various definitions often focus on big data as a collection of huge volumes of unstructured data (Amado et al. 2018), alternatively it is considered as a collection of data being taken from different sources and integrated for another use (for example using administrative data). It is not a well-defined tangible object, and the potential of using Big Data to solve problems depend on what the problem is, what sources of Big Data may contribute to the solution, and whether any inherent biases or measurement errors with those sources are significant enough to make them unsuitable for the solution (Tam and Clarke 2015). Big Data is often defined by its characteristics along four dimensions (Daas and Puts 2014; Morabito 2015):

- volume—the number of data records, their attributes and linkages;
- velocity—how fast data are produced and changed, and the speed at which they must be received, processed and understood;
- variety—the diversity of data sources, formats, media and content;
- veracity—quality of the data and trust in the data.

Because of these discrepancies with definitions there is a need to investigate how the users of the data view the usefulness, and whether businesses are capable of exploiting these opportunities that big data offers. While plenty of scholars are conducting research on Big Data with respect to Tourism marketing applications, less is found in addressing specifically the benefits that marketers could potentially achieve through Big Data solutions. While Big Data adoption within the industry exists, more research is needed to clearly identify the pros and cons for organizations to invest in Big Data (Amado et al. 2018). There are clear foci of organizations to adopt Big Data to extract strategic value but the data-driven decision making is still in its infancy (McAfee and Brynjolfsson 2012; Russom 2013). It is better to focus on decisions based on evidence from sources such as data, than to use intuition alone. Big data can provide better evidence and insights to support the process of decision making.

11.5 Tourism Marketing and Big Data

The enormous growth of powerful digital devices and relevant software create opportunities for collecting and leveraging voluminous data and offer ways for tourism and hospitality firms to innovate and create better offerings and evaluation for better performance (Moro et al. 2017).

The effectiveness of a Big data system heavily depends on its set up for the extraction and synthesis of information packets to support business decisions, such as the types of markets and segments to target, what mix of marketing communications to use, and how to shape the characteristics of the offering (Grossman and Siegel 2014). Also, the rest of tourism/hospitality marketing mix offering, i.e. characteristics, price, and channels of distribution, can be formulated to better match potential tourists' profiles (Hazen et al. 2014). Predictive analytics utilizing Big Data drawn specifically for Marketing purposes take the form of Marketing analytics solutions that implement Marketing underpinnings and techniques to solve real-world Marketing problems (Grigsby 2015).

Among others, Malthouse et al. (2013) proposed that traditional customer relationship management (CRM) can greatly benefit from the Big Data evolution and they proposed specific guidelines accordingly. The predictive methods are useful to provide snapshots of the CRM data to extract models for predicting success of future campaigns and offer guidance to managers to make informed strategic decision-making about their business (Brynjolfsson et al. 2011). Other researchers have also identified the need to incorporate analytics in CRM encompassing widely known concepts in Marketing, such as the evaluation of customer lifetime value (Moro et al.

2015). These impacts are seen as creating real value from a marketing perspective because traditional data analyses methods have not been able to provide information on big volumes of data, such as text, pictures and web data.

The focus now is on leveraging the abundant data available from personal smart devices, e.g. smartphones, smartwatches, making this data a source of information reflecting individuals' preferences, lifestyles and habits. Recent studies confirm that while some work has been done to fill this gap, the size of the gap is increasing at the velocity of the inflation of Big Data availability as more individuals engage with smart devices (e.g., Erevelles et al. 2016). Thus, relevant opportunities for producing tourism marketing applications (tour-apps) based on full scale usage of Big Data will likely grow in the future.

11.6 Considering Visitors' Needs

The hotel industry is a highly competitive one in that hotel firms offer essentially homogeneous products and services. The competition, thus, drives the desire of hotels to differentiate themselves among their competitors. Guest satisfaction and approval has become one of the key measures of a hotel's ability to outperform others. Understanding the needs of customers and their experiences is vital for improved sales and overall organization performance. Recent studies have used big data analytics to better understand important hospitality issues, e.g. the relationship between hotel guest experience and satisfaction. One such study by Xiang et al. (2015) found several dimensions of guest experiences that carried varying weights and, more significantly, meaningful semantic compositions. The association between guest experiences and satisfaction was strong, suggesting that these two domains of consumer behavior are fundamentally associated.

Banerjee and Chua (2016) have collected and analyzed reviews of customers taken from TripAdvisor. They found that travelers' rating patterns differed significantly depending on the chain versus independent hotels. Ye et al. (2011), as well as Lu et al. (2014) found a significant impact of user-generated reviews on hotel sales. Liu et al. (2017) investigated a collection of 412,784 user-generated reviews on TripAdvisor for 10,149 hotels from five Chinese cities. They found that overseas tourists, who speak diverse languages (English, German, French, Italian, Portuguese, Spanish, Japanese, and Russian), vary significantly on the roles of various hotel attributes ("Rooms," "Location," "Cleanliness," "Service," and "Value") and in providing their overall satisfaction rating for hotels. Chinese tourists domestically, prefer room-related hotel attributes compared to overseas tourists. From prior studies on the issue, user-generated reviews yield a reliable source of information that can be used to understand the drivers of hotel satisfaction and provide feedback to meet any future customer service demand. The use of these types of feedback require organizational processes for understanding and responding to the customer needs. This might better be considered from internal and external frameworks for tourism and hospitality organizations.

11.7 Big Data Frameworks for Tourism and Hospitality

11.7.1 Big Data Flows of a Tourism and Hospitality Organization

For tourism and hospitality organizations to be analytical and use Big data approaches, a specific set of frameworks are needed to enable the organizations to participate in the Big data evolution. As these organizations contribute to the formation of the tourism service ecosystem, and further to the S-D logic introduced by Vargo and Lusch (2004) illustrates the dynamic interactions between various industry stakeholders such as the Viable Systems Approach (VSA) (Golinelli et al. 2012; Barile et al. 2012). The VSA is particularly suitable for building a theoretical framework in this field of study, as it can support the understanding of the organizational and ecosystem change introduced due to the Big data evolution.

In this vein, we have developed two frameworks that support better understanding and potential implementation of Big data processes for tourism-related organizations. The first framework depicted on Fig. 11.1 has the tourism/hospitality organization at the epicenter. It demonstrates the interrelationships and flows of data between the organization and other business or institutional stakeholders.

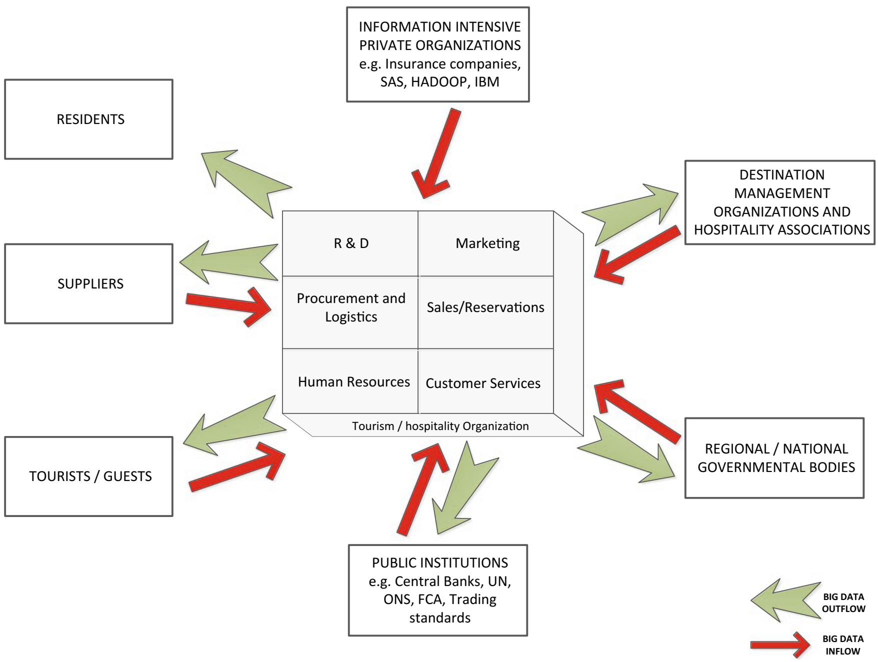


Fig. 11.1 Big data flows of a tourism and hospitality organization (BDF). *Source* The authors

The cuboid shows the various departments within the organization. Typical examples of departments are included here including the research and development (R&D), marketing, procurement and logistics, sales/reservations, human resources and customer services. The internal flows show the data flowing into the organizations as represented by red arrows. The big data for the outflow are presented with green arrows. An example of an inflow will be suppliers who provide products or services to the organization. A Big data application might focus on the inflow of big data which has been analyzed by applying predictive analytics from volumes of data based on recurring patterns of stock provide richer information to provide a customized service based on the type of guests or customers in the tourism or hospitality organization. For example, customers/guests fitting a particular profile are presented with offerings that are based on their requirements from previous spending or hotel visits.

On the other hand, outflows of data might focus on the big data collected and processed by public institutions, such as the data required to provide data on customer visits, rooms booked and or payments made to hospitality organizations. This data is made available on an aggregate basis to government organizations such as the Office for National Statistics (ONS) for official statistics on quarterly hotel visits, overnight stays and relevant payments made. These big data outflows provide useful information from the organization to provide global information about data trends and patterns for the benefit of the industry. This type of data provides useful indicators about the health of the whole industry, and thus benefits multiple parties. Subsequently, once these data have been collected and processed by the public institutions, they flow in as feedback (e.g. tables, media release, general reports) to all tourism/hospitality organizations. Similar inflows and outflows can be attributed to; residents, suppliers, information intensive private organizations, governmental bodies and, DMOs and hospitality industry associations and the tourism/hospitality unit.

Table 11.1 below has been developed to show how these inflows and outflows work in practice between the tourism/hospitality organization and external organizations, guests, tourists and residents. It illustrates various types of data inflows then outflows into tourism/hospitality organizations, examples of the typical data flows, resources that are required to action these transfers/informatization. Lastly, it considers whether the necessary enterprise-wide organizational processes, individual skills and expertise, and interrelationships between stakeholders are emerging or established capabilities in the tourism ecosystem.

11.7.2 The Big Data Tourism Analytical (BDTA) Framework

The BDF framework that is depicted as a satellite system of information carriers that orbit around a tourism/hospitality organization, shows the data that flows in and out of the organizational boundaries. Nevertheless, it does not illustrate the integration of the context and the wider interactions that are pertinent within a Big data analytical environment. Since, data storage is only one part of big data analytics, we should

Table 11.1 Big data flows of tourism and hospitality organisations

Business or Institutional Stakeholders	Specific examples of Organisations and related activities	Big data process resources	Established/emergent capabilities
Information intensive private organisations to the tourism organisation	Insurance companies' data for determining liability for the destination management events	Data analysts/database administrators	Emergent—in respect of the use of Big data Brown et al. (2011), Manyika et al. (2011)
Destination management organisations	Data about sales and marketing are exchanged with the DMOs and/hospitality associations and aggregated to provide a view of the industry and the hospitality organisation	Databases, software to share data	Established based on the industry view according to Kunz et al. (2017), Zhang (2012)
Regional/national governmental bodies	Municipalities and counties providing reports, research, strategic data on regional or national industry aspects and sharing with the hospitality organisation	Expertise of Insights analysts, reports, research on regional or national outcomes/goals. Databases, software for updates etc	Emergent—in the sense of using data to create value Beer (2018), Gretzel et al. (2015b)
Public institutions	Sharing of data between the Office of National Statistics (ONS) and the DMO about customer visits, sales transaction data. Could be based on compliance for tax or for official statistics	Database or CRM capturing and sharing interval data on visitors, sales transactions	Established—this practice is well established according to Beer (2018) who explains the value of these partnerships
Tourists/Guests	Reviews of previous guest/tourist experiences of DMOs based on feedback on sites such as Tripadvisor, Booking.com, Trivago etc	Online reviewing uses Big data and ranks feedback from Reviews posted about guests/tourists' experiences. DMOs need to engage with feedback	Emergent—based on research (see Buhalis and Foerste 2015; Gretzel et al. 2015b; Xiang et al. 2015)

(continued)

Table 11.1 (continued)

Business or Institutional Stakeholders	Specific examples of Organisations and related activities	Big data process resources	Established/emergent capabilities
Suppliers	RFID tags are used to track and trace products such as towels for monitoring their location, frequency of cleaning and or whether they have been stolen	RFID tags, software for data analytics, track and trace geotags, databases etc	Emergent—there are new ways of working with suppliers based on reporting and data tracking etc. Golinelli et al. (2012), Barile et al. (2012), Priporas et al. (Priporas et al. 2017)
Residents	People living locally are provided information about the operations, processes of the organisation that affects them. For example, unstructured data from surveillance, security etc. to ensure safety and security is monitored	Camera footage and monitors are used for surveillance and data is stored etc	Established—the practice is not new but the technology and dynamic data integration with websites and other device is developing (e.g., Erevelles et al. 2016)

consider the wider interactions of how intelligence is accessed by various stakeholders in the tourism ecosystem. This is where Big data emerges, i.e. to facilitate the multiple exchanges and processes at various layers of data infrastructure.

Figure 11.2 below provides an overview of the big data flows that cross over and processed by the tourism analytical hub on a real-time basis. The framework extends across four layers of data flows from a series of access points into the analytical hub. The analytical hub is where the Big data is transformed and pumped throughout the ecosystem. These layers include, firstly, the access points serving as data satellites; secondly, types of data used by the access point; thirdly, analytical professionals required for analyzing the big data; and lastly, in the center is the tourism analytical hub.

The **first layer** presented in the framework spans nine key access points. These are commercial banks, insurance organizations, governmental institutions, social networking platforms, tourism services vendors, tourists/guests, destination management organizations, tourism/hospitality organizations, and the supply chain for both goods and services. The data from these organizations provide big data to hospitality and tourist organizations to deliver better touristic services. For example, *commercial banks* (access point 1) offer data on transactions and electronic payments, thus providing financial intelligence which is processed and analyzed in the tourism analytical hub in real-time. The big data that flows from the commercial banks are transformed by processes such as predictive analytics to provide scoring reports, profiling of customers and credit history to provide better tailored services and better insights about tourists.

Access point 4 focuses on the social networking platforms and the big data that provides information about the feedback of customers and guests. The Big data provide information about the customer satisfaction and information about their experiences in hotels and tourist organizations. It is important to have a way to analyze the data and information presented on social media. With the plethora of data provided from these social media platforms it is in the best interests of the hospitality and tourist organizations to consider the feedback and reviews from customers on the platforms. The ability to analyze the Big data is possible with the advent of key data consultants able to react to feedback swiftly and also detect when there are anomalies in the feedback patterns on social platforms. When data is received, text analytics software can decipher when feedback is being fabricated. This gives the whole process higher credibility, authenticity and makes the reviews and other information shared via the tourism platforms more trustworthy (e.g. Tripadvisor, Booking.com). Furthermore, the social media feedback and reviews are useful to capture how employees and the organization is performing. Within the tourist analytical hub there are now ways to integrate Big data and provide feedback across multiple social media sites.

Access point 7 focuses on the Big data flows of *destination management organizations* (*DMOs*). The Big data flows show that reporting, forecasting and insights are some of the important advantages of the data for these organizations. External data from the weather are modelled and insights are provided about the potential weather system that might impact on the destination management organizations performance. Diving organizations are able to get updates on weather models, sea conditions and global positioning systems for key information affecting their organizations. Updates on storms, tides, hurricanes and other weather systems can have a significant impact on these organizations ability to deliver positive customer experiences such as the ability to experience diving at key destinations. It would not be pleasurable with murky, choppy seas and, potentially, safety could be compromised. With the advent of Big data, integrated solutions benefit these organizations, especially since they are dependent on external data, as well as internal business intelligence to thrive and prosper.

The tourism/hospitality organization represents access point 8 which shows the flow of big data into the tourism analytical hub. For example, ticketing information for events or hotel visits are able to be processed in real-time from external as well as internal data. The data flow from the computer terminal interfacing with the website for ticketing big data will access the tourism analytical hub. The CRM in the hub is connected to the Big data and the Cloud and provide feedback on the availability of tickets or rooms in the hotel. For tourism and hospitality organizations, the advantages are that with the Big data efficiency and effectiveness of service offerings.

Figure 11.2 provides additional details about the rest of access data points/stakeholders contributing to the function of the tourism ecosystem from an intelligence point of view.

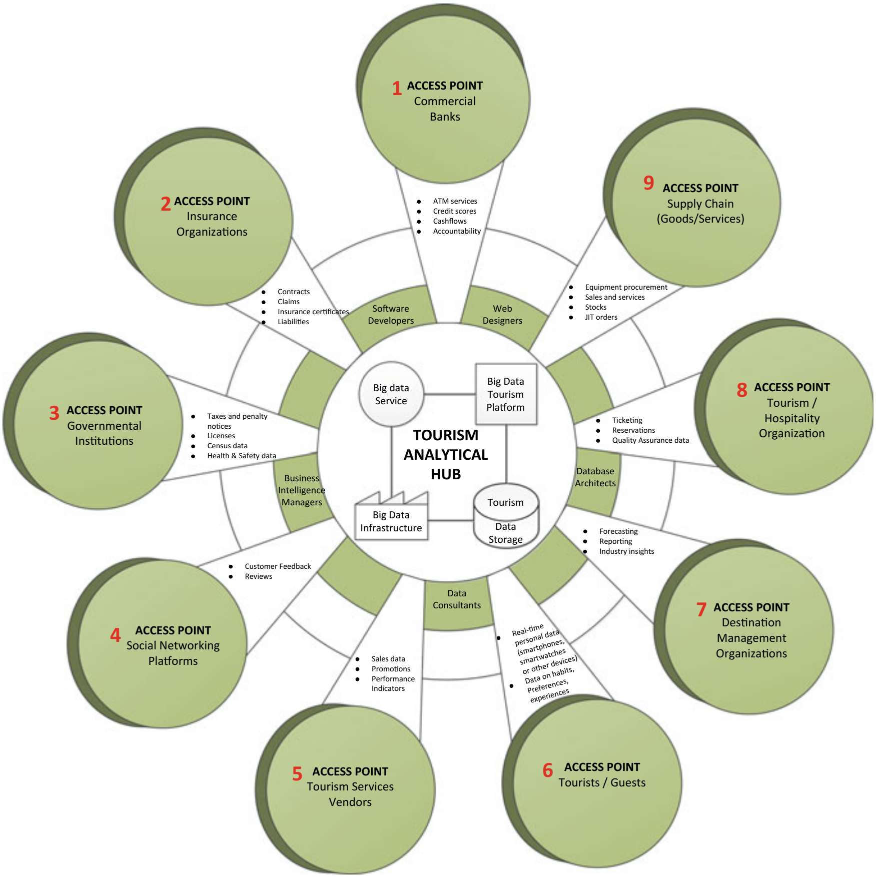


Fig. 11.2 The Big data tourism analytical (BDTA) framework. *Source* The authors

The **second layer** from the analytical hub focuses on professionals who have expertise in providing analytical services for transforming the data into key business intelligence. Data analysts, such as the software developers, web developers, data architects and data analysts, use business intelligence by transforming the data. They provide the key analytics capability to convert data into useful—and potentially optimum—decisions within organizations and, thus, they are paramount to shaping strategies and the future performance of organizations. For the analytics environment to thrive, transformed big data is needed to provide better decisions to the ecosystem.

The **third layer** represents the spokes which illustrate the specific types of data between the access points of data and the analytical hub. Without the oil of data provided both internally and externally to the organizations, the tourism and hospitality ecosystem is limited in their ability to be effective and efficient. Big Data opportunities transform into long-lasting success only for organizations focused on managing synergy between a firm's execution of engagement efforts and the customers' experience, motivation, preferences and expectations (Kunz et al. 2017).

The **fourth layer** is the tourism analytical hub which provides the infrastructure for doing the analytics. The tourism analytical hub is a part of the Cloud (Buhalis and Amaranggana 2013) and the 'heart' of the wider analytical ecosystem. The tourism analytical hub includes *the big data service, which* provides the data which could be collected from internally or externally to the organizations; the *big data platform,* such as Integration Platform as a Service (iPaaS) and Blockchain or other proprietary software to support Big data processing and accessing without installing any hardware devices or middleware; and *big data storage,* such as the data marts, and servers. *Big data infrastructure* provides the interfaces between the CRM or proprietary software and software which communicates with the servers such SQL server and the Database Management System (DBMS). The tourism analytical hub provides the engine for the ecosystem to survive and thrive. Ultimately, data is needed to service the engine and provide smooth performance for the hospitality and tourism organizations. Without the hub the ability for the analytics is limited, as Big data and the predictive and forecasting techniques, will be stifled. Much as the engine needs oil to ensure that the engine is lubricated, similarly the tourism analytical hub needs volumes of Big data to create supportive information for decision making. These decisions can now be made real-time via analyzing feedback provided by various stakeholders. Thus, the tourism analytical hub is needed to provide infrastructure for predicting how to improve services and feedback on performance. It is the networking power it carries that fights silo-ed intelligence practices and solves problems holistically. Big data is a major game changer now in optimizing change and better managing future events.

Table 11.2 focuses on the roles, key skills required, level of competencies required, uses of big data and the extent to which the industry is ready to embrace the challenges and opportunities afforded by big data for better business intelligence within the tourism sector.

11.8 Conclusion

This chapter highlighted the active role Big data take in shaping the contemporary tourism and hospitality industry intelligence. To further understand this, approaches to implementing Big data were explored within a tourism and hospitality context. Two frameworks were developed, providing a blue print showing the various stages of Big data, thus demonstrating the relevant flows and implementation processes. Marketing analytic solutions utilizing Big data may benefit tourism organizations in achieving real-time deep insights into tourists' preferences (Moro et al. 2014), constructing interactive reports and dashboards for managers or even unveiling interesting trends from what is being said about the tourism destination or hospitality organization on social media (Miah et al. 2017; Morabito 2015). Furthermore, big data-driven marketing practices, such as recommendations, geo-fencing, search engine marketing,

Table 11.2 Key skills, competencies, and uses of big data

Actor/Stakeholder	Key Skills required	Level of competency required	How big data used	Resource availability
Software developers	Programming skills Analytical skills Business intelligence Analytics product knowledge e.g. Hadoop, Java, SQL, Machine learning frameworks, team work skills	Highly competent in programming languages and software such as Hadoop and Java	Developers code programmes that communicate between different databases and websites. For example, they create software for dataflows between the booking system and the payment system	Currently, there is a shortage of key skilled developers with this set of skills, especially with AI and Hadoop skills
Web designers	Programming skills Analytical skills Business intelligence Analytics product knowledge e.g. Hadoop, Java, SQL, team work skills	Competent at developing websites that work with data integration in CRM, reporting tools etc	Web designers create and code the website details used for systems such as the CRM, organisation websites	There are shortages of web designers with experience of dealing with Big data web applications/analytics
Database architects	Programming skills Analytical skills Business intelligence Database Infrastructure knowledge, Analytics product knowledge e.g. Hadoop, Java, SQL, team work skills Conceptual understanding of the enterprise data, products, applications and infrastructure	Highly competent at providing the infrastructure to support existing and enterprise-wide hardware and software for application in the organisation	They are designers of the overarching architecture for hardware and software data flows with the organisation	There is a shortage of database architects with skills in developing key infrastructure to support an enterprise wide Big data strategy. With the increase in demand for these skills the market is not able to keep pace with the demand

(continued)

Table 11.2 (continued)

Actor/Stakeholder	Key Skills required	Level of competency required	How big data used	Resource availability
Data consultants/analysts	Analytical skills Business intelligence Advanced data analyses skills Analytics product knowledge e.g. Hadoop, SQL, SAS	Specialised knowledge is required to analyse the data for reporting, analyses and maintaining the integrity of the data	Data consultants and analysts are key in providing meaningful transformation of data for practical decision-making and evidence. They provide the data for making meaningful business insights, cleaning, preparing, analysing and ongoing reporting of key metrics	These are key roles in organisations to make sense of the data. They are in high demand for their expertise and they are definitely needed to ensure the meaning is created from the volumes of data
Business Intelligence Managers	Key knowledge of business needs—business subject matter expertise Business intelligence, analytics, Decisions based on data Management expertise	High levels of competence is required to understand how to make the best use of the data and business models for business intelligence	They take the Big data and use their subject matter expertise to make better decisions using the data. For example, they are able to use reservation information for strategic planning and predictions of reservations for customer visits etc	Making decisions based on data, requires specialised knowledge and expertise in this area. There are shortages of the personnel to manage analysts and provide business intelligence to the organisation
Business Analysts	Analytical skills Reporting skills Documenting user requirements, Developing process flow diagrams Business intelligence, team work skills	Business analysts need to be competent in their understanding of requirements of business and IT professionals	They provide the detailed documentation and communication about the processes involved between the subject matter experts and the IT personnel	Bridging the difference between IT and business areas, the Business analysts are important to provide the communication of user requirements and mapping the processes for IT to develop

CRM, market segmentation, customization and marketing mix optimization, play a key role in creating new forms of data-driven strategies and enabling business innovation (Priporas et al. 2017).

There is excitement and keen anticipation that future Big data applications will provide optimized services for organizations in tourism and hospitality sector when they leverage the opportunities arising from technological advances. The game change is about the integration of business intelligence using Big data occurring throughout the tourism ecosystem. Decision making capabilities may be much improved to the benefit of the whole performance (or micro) environment of a tourism/hospitality organization. Future research could investigate empirically the Big data flows across the tourism industry by utilizing the proposed BDTA framework. This would offer insights into how Big data influences operations of each organization involved, as well as the performance of the tourism service ecosystem as a whole. In this regard, researchers would be able to suggest improvements for the way access points are efficient and effective; consequently, tourism and hospitality managers could draw on this research and inform their practice.

Big data calls for collaborative action. Therefore, we suggest: "*Think holistically; think Big data for more intelligent decision-making, supreme guest experiences and enhanced business performance. The Big data era is here. The game has changed indeed*."

References

Amado A, Cortez P, Rita P, Moro S (2018) Research trends on big data in marketing: a text mining and topic modeling based literature analysis. Euro Res Manage Business Economics 24(1):1–7

Banerjee S, Chua AY (2016) In search of patterns among travellers' hotel ratings in TripAdvisor. Tourism Manage 53:125–131

Barile S, Pels J, Polese F, Saviano M (2012) An introduction to the viable systems approach and its contribution to marketing. J Business Market Manage 5(2):54–78

Bean R, Kiron D (2013) Organizational alignment is key to big data success. MIT Sloan Manage Rev 54(3):1–6

Beer D (2018) Envisioning the power of data analytics. Information Commun Soc 21(3):465–479

Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. Science 323(5916):892–895

Brown B, Chui M, Manyika J (2011) Are you ready for the era of 'big data'. McKinsey Q 4(1):24–35

Brynjolfsson E, Hitt LM, Kim HH (2011) Strength in numbers: how does data-driven decision making affect firm performance? Social Science Research Network (SSRN) [Online]. Available at https://ssrn.com/abstract=1819486 or http://dx.doi.org/10.2139/ssrn.1819486.

Buhalis D, Amaranggana A (2013) Smart tourism destinations. Information and communication technologies in tourism 2014. Springer, Cham, pp 553–564

Buhalis D, Foerste M (2015) SoCoMo marketing for travel and tourism: Empowering co-creation of value. J Destination Marketing Manage 4(3):151–161

Daas PJ, Puts M (2014) Big Data as a source of statistical information. Survey Statistician 69(1):22–31

Davenport TH, Harris JG (2007) Competing on analytics: the new science of winning. Harvard Business Press

Dedić N, Stanier C (2016) Measuring the success of changes to existing business intelligence solutions to improve business intelligence reporting. In: Tjoa AM, Xu LD, Raffai M, Novak NM (Eds), Research and practical issues of enterprise information systems. 10th IFIP WG 8.9 Working Conference, CONFENIS 2016, Vienna, Austria, December 13–14, 2016, Proceedings. Springer, Cham: pp 225–236

Erevelles S, Fukawa N, Swayne L (2016) Big Data consumer analytics and the transformation of marketing. J Business Res 69(2):897–904

Ferretti V, Comino E (2015) An integrated framework to assess complex cultural and natural heritage systems with Multi-attribute value theory. J Cultural Heritage 16(5):688–697

Fuchs M, Höpken W, Lexhagen M (2014) Big data analytics for knowledge generation in tourism destinations–a case from Sweden. J Destination Marketing Manage 3(4):198–209

Goldenberg J, Libai B, Muller E (2001) Talk of the network: a complex systems look at the underlying process of word-of-mouth. Marketing Lett 12(3):211–223

Golinelli GM, Barile S, Saviano M, Polese F (2012) Perspective shifts in marketing: toward a paradigm change? Ser Sci 4(2):121–134

Gretzel U, Sigala M, Xiang Z, Koo C (2015a) Smart tourism: Foundations and developments. Electronic Markets 25(3):179–188

Gretzel U, Werthner H, Koo C, Lamsfus C (2015b) Conceptual foundations for understanding smart tourism ecosystems. Comput Human Beha 50:558–563

Grigsby M (2015) Marketing analytics: A practical guide to real marketing science. Kogan Page Publishers, London

Grossman RL, Siegel KP (2014) Organizational models for Big Data and analytics. J Organisational Design 3(1):20–25

Guo Y, Liu H, Chai Y (2014) The embedding convergence of smart cities and tourism internet of things in China: an advance perspective. Adv Hospitality Tourism Res 2(1):54–69

Hazen BT, Boone CA, Ezell JD, Jones-Farmer LA (2014) Data quality for data science, predictive analytics, and big data in supply chain management: an introduction to the problem and suggestions for research and applications. Int J Prod Economics 154:72–80

Kunz W, Aksoy L, Bart Y, Heinonen K, Kabadayi S, Ordenes FV, Sigala M, Diaz D, Theodoulidis B (2017) Customer engagement in a big data world. J Serv Marketing 31(2):161–171

LaValle S, Hopkins MS, Lesser E, Shockley R, Kruschwitz N (2010) Analytics: the new path to value. MIT Sloan Manage Rev 52(1):1–25

Liu Y, Teichert T, Rossi M, Li H, Hu F (2017) Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. Tourism Manage 59:554–563

Lu Q, Ye Q, Law R (2014) Moderating effects of product heterogeneity between online word-of-mouth and hotel sales. J Electron Commer Re 15(1):1–12

Maglio PP, Spohrer J (2008) Fundamentals of service science. J Academy Marketing Sci 36(1):18–20

Malthouse EC, Haenlein M, Skiera B, Wege E, Zhang M (2013) Managing customer relationships in the social media era: introducing the social CRM house. J Interactive Marketing 27(4):270–280

Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C et al (2011) Big Data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Marine-Roig E, Clavé SA (2015) Tourism analytics with massive user-generated content: A case study of Barcelona. J Destination Marketing Manage 4(3):162–172

McAfee A, Brynjolfsson E, Davenport TH, Patil DJ, Barton D (2012) Big data: the management revolution. Harvard Bus Rev 90(10):60–68

Miah SJ, Vu HQ, Gammack J, McGrath M (2017) A big data analytics method for tourist behaviour analysis. Information Manage 54(6):771–785

Morabito V (2015) Big data and analytics. Strategic and organisational impacts. Springer, Cham

Moro S, Rita P, Coelho J (2017) Stripping customers' feedback on hotels through data mining: the case of Las Vegas Strip. Tourism Manage Perspect 23:41–52

Pantano E, Priporas CV, Stylos N (2017) 'You will like it!' Using open data to predict tourists' response to a tourist attraction. Tourism Manage 60:430–438

Phillips-Wren G, Hoskisson A (2015) An analytical journey towards big data. J Decisi Syst 24(1):87–102

Priporas CV, Stylos N, Fotiadis AK (2017) Generation Z consumers' expectations of interactions in smart retailing: A future agenda. Comput Human Beha 77:374–381

Russom P (2013) Managing big data. TDWI Best Practices Report, TDWI Res 1–40

Tam SM, Clarke F (2015) Big data, official statistics and some initiatives by the Australian Bureau of Statistics. Int Stat Rev 83(3):436–448

Vargo SL, Lusch RF (2004) Evolving to a new dominant logic for marketing. J Marketing 68(1):1–17

Xiang Z, Schwartz Z, Gerdes JH Jr, Uysal M (2015) What can big data and text analytics tell us about hotel guest experience and satisfaction? Int J Hospitality Manage 44:120–130

Ye Q, Law R, Gu B, Chen W (2011) The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. Comput Hum Behav 27(2):634–639

Zhang L (2012) Smart tourism: the coming age of customization and intelligent public services. J Tourism Tribune 27(2):3–5

Chapter 12
Strengthening Relational Ties and Building Loyalty Through Relational Innovation and Technology: Evidence from Spanish Hotel Guests

Irene Gil-Saura, María-Eugenia Ruiz-Molina and David Servera-Francés

Abstract Smart tourism relies on the ability to collect enormous amounts of data and intelligently store, process, combine, analyse and use big data to inform business innovation, operations and services. Companies are becoming aware of the potential of analytics and big data and the need to invest in technology to remain competitive. Most of the research on big data, technology and innovation in tourism focuses on the firm's perspective, paying scanty attention to customers' perceptions. The present chapter aims to analyse the impact of hotel relational innovation and technology on brand equity and hotel-guest relational ties and customer loyalty. This objective has been pursued in quantitative research through personal interviews with Spanish guests of 3, 4 and 5-star hotels in the Valencia region, one of the top holiday destinations in Spain. The sample of 401 guests at 42 hotels provides evidence to suggest a significant positive impact of relational innovation on guest perceptions of hotel technology and strength of relational ties. Moreover, information and communication technologies exert a positive impact on overall brand equity which, in turn, has a positive impact on relational ties and guest loyalty. Relational innovation makes a particularly strong contribution to the strength of relational ties and ultimately, loyalty. Empirical research is urgently needed in this area to gain insight into these relations and the role of big data in value creation.

Keywords Relational innovation · Information and communication technology · Big data · Brand equity · Strength of relational ties · Loyalty · Hotels

I. Gil-Saura · M.-E. Ruiz-Molina (✉)
Universitat de València, Valencia, Spain
e-mail: m.eugenia.ruiz@uv.es

I. Gil-Saura
e-mail: irene.gil@uv.es

D. Servera-Francés
Universidad Católica de Valencia, Valencia, Spain
e-mail: david.servera@ucv.es

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_12

12.1 Introduction

The service industry is a highly innovative sector (Oke 2007). The interrelated nature of innovation in services means that change in one dimension (e.g. emergence of a new technology) necessarily affects other aspects of the service or leads to new knowledge, skills and processes (Den Hertog 2000), hence the interest in researching innovative initiatives (Ganesan et al. 2009; Hu and Zhao 2010; Ritter and Walter 2012; Sansone and Colamatteo 2017).

Researchers have shown a special interest in relational innovation in distribution channels (Ganesan et al. 2009), but not in service industries like tourism. Several works on tourism have closely associated technology with innovation (e.g. Orfila-Sintes et al. 2005), but usually analyse the hotel manager's perspective and ignore the customer's point of view. Moreover, analysis of the degree of innovation and technological advancement commonly adopts a merely descriptive approach.

The literature emphasises that hoteliers can achieve greater differentiation by offering the latest technology to guests, enhancing their experience and attracting new customers, with the potential to increase revenues (Brochado et al. 2016). Previous research demonstrates the utility of big data analytics to better understand the relationship between the hotel guest experience and satisfaction (Xiang et al. 2015), a long-standing topic of interest because of its widely recognized contribution to customer loyalty, repeat purchases and positive word-of-mouth (Oh and Parks 1997).

There is evidence to support the notion that businesses that can harness data analytics significantly outperform their competitors in business performance (He et al. 2016). However, although big data analytics does not preclude hypothesis testing, it is often applied to explore novel patterns or predict future trends from the data (Aiden and Michel 2014). According to Morabito (2015), operating in a big data driven, smart environment affects all nine elements of business models: (1) customer segments; (2) value propositions; (3) channels; (4) customer relationships; (5) revenue streams; (6) key resources; (7) key activities; (8) key partnerships; and, (9) cost structure. Focusing on customer relationships, the mechanisms through which guest perceptions of hotel technology may influence the outcomes of the establishment are still a rather underexplored topic.

Several researchers note that it is vital to understand the changes taking place in the tourist experience as well as the role of technology in the process (Neuhofer and Buhalis 2012). For instance, Lam et al. (2017, p. 22) include in the research agenda to analyse the combination of big data and "small data" generated by frontline employees to answer this question: "How do customers react to "supersmart" hard- and software solutions that substitute "human" frontline employees?". However, to the best of our knowledge, the role of relational innovation and technology in brand equity creation and its outcomes in hospitality has not been investigated. The present paper aims to examine the connections between relational innovation, hotel degree of technological advancement, brand equity, strength of relational ties and guest loyalty towards the hotel from the consumer perspective. We attempt to infer relevant theoretical and managerial implications regarding the contribution of relational innovation and technology to hotel brand equity and strengthen the links of the hotel guest with the establishment.

In pursuit of this aim, in Section "12.2" we review the literature on relational innovation and we attempt to ascertain the link between this construct and big data in tourism. Next, in a similar vein, Section "12.3" discusses the relevance of Information and Communication Technologies (ICT) in general, and big data in particular, in the tourism industry. Then the hypotheses are developed and a model gathering the hypothesized relations is presented in Section "12.4". Sections "12.5" and "12.6" describe the method and the quantitative results, respectively and the Conclusions close this chapter.

12.2 Relational Innovation and Big Data in Tourism

In a smart tourism ecosystem, actors may possess and exchange tangible and intangible resources (e.g., tools, software, and information); human resources (e.g., skills, knowledge, and virtual communities); and relational resources (e.g., relations to partners and suppliers, and network membership). All stakeholders are actors aiming to interact and exchange resources with other actors for value co-creation (Gretzel et al. 2015). Traditional labels and roles assigned to players like travellers, firms and intermediaries are no longer valid (Vargo and Lusch 2008), since any type of stakeholder can become a producer, consumer, intermediary and so on, depending on resources and connections (Gretzel et al. 2015). Therefore, producer-client relationships must be redefined and new approaches developed to cooperation in production, delivery and consumption of services (Anttiroiko et al. 2014). In this scenario, a research line has emerged in recent years focusing its attention on the concept of "relationship innovation" or "relationship-based innovation" (Ganesan et al. 2009), based on improving relations between the company and its customers (Dupuis 2002; Lin 2015) as well as its suppliers (Drejer 2004; Ruiz-Molina et al. 2017). Similarly, "relational innovation" has been used to refer to "new ways and methods for buyer-seller interactions in the market" (Dantuma and Hawkins 2001, p. 89).

Innovation refers to the fulfilment of novel ideas held within an organization (Amabile 1983; Wang and Tsai 2014). However, there is no consensus on the content and dimensionality of this construct. Concerning relationships in the value chain, the United Nations Industrial Development Organization points out four main approaches to understanding innovation, that is, product innovations, process innovations, functional innovation and innovation in the value chain (UNIDO 2002).

In the academic context, there have been attempts to go deeper into the identification of innovation initiatives through relations between members of the distribution channel (Ganesan et al. 2009; Hu and Zhao 2010). For instance, Djellal and Gallouj (2010) consider three main dimensions of product/service innovation, process innovation and external relational innovation. Alternatively, Musso (2010) states that innovation in marketing channels may be analysed from three different approaches, namely the technological perspective, relational perspective and structural perspective. The technological approach to relations with the end consumer comprises a series of information and communications technology applications focused on cash management, electronic and mobile payment systems, checkout technologies, dynamic pricing policies, payment methods, distance and on-line sales, and self-service technologies, among others (Musso 2010). The relational perspective refers to applications that are not innovative in themselves, insofar as they have already been developed and diffused widely, but they represent areas where firms can develop innovative solutions such as an assortment of ethical and socially responsible products, loyalty programmes and initiatives designed to guarantee the firm's environmental sustainability (Ganesan et al. 2009; Hu and Zhao 2010). According to the relational perspective, innovations in marketing channels include customer care initiatives (e.g. Customer Relationship Management), among others (Musso 2010). Finally, the structural approach focuses on new configurations of channels, new channels or multiple channels introduced either on the initiative of industrial firms (e.g. franchises, multilevel direct channels, electronic commerce) or as the result of internal mechanisms (e.g. multichannel strategy).

In tourism, although innovation has been traditionally analysed from a technological perspective (e.g. Orfila-Sintes et al. 2005; Guttentag 2010; Smerecnik and Andersen 2011; Neuhofer et al. 2014), several studies argue that innovation goes beyond technology, adopting a wider approach that gathers new marketing ideas and new practices that are profitable for the company (e.g. Jacob et al. 2003; Weerawardena 2003; Novelli et al. 2006; Ottenbacher et al. 2006; Pikkemaat and Peters 2006; Volo 2006; Sundbo et al. 2007; Martínez-Ros and Orfila-Sintes 2009; Vila et al. 2012; Pivčević and Garbin Praničević 2012; Tejada and Moreno 2013; Krizaj et al. 2014; Mattsson and Orfila-Sintes 2014).

In highly competitive environments like the hospitality industry where hotel brands offer essentially homogeneous products and services, firms need to distinguish themselves from their competitors (Xiang et al. 2015). Innovation enables hotels to adopt novel ideas, improve service processes, enhance operational efficiency, meet customer needs, and maintain competitive advantages in rapidly-changing markets (Grissemann et al. 2013; Nicolau and Santa-María 2013).

To develop long-term relationships with customers, many firms have invested in data-intensive customer relationship management (CRM) systems (Mendoza et al. 2007). CRM has been defined as a strategy for acquiring, enhancing the profitability of, and retaining customers, enabled by a technological application to collect, store and analyse data for achieving mutual benefits for both the organisation and the customers (Rababah et al. 2011). The CRM process may be associated to the use of loyalty cards that allow the company to obtain a great deal of information from its customers (Musso 2010). Loyalty card technology allows companies to transform data on consumer behaviour into 'learning' relationships (Pine et al. 2009) and customer loyalty (Mauri 2003). Indeed, all companies use data in some way to conduct business (Hartmann et al. 2016), and data are available to companies from internal and external sources in a variety of formats, that is, structured data such as transactional data and unstructured data such as customer comments on social media (Phillips-Wren and Hoskisson 2015). "Big data", defined as "high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" (Gartner 2012), has led to new opportunities but also challenges for companies to develop analytical capabilities for existing data and integrate information from these new data sources into their CRM decisions (Phillips-Wren and Hoskisson 2015). Laney (2001) describes "big data" technologies as a new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data, and points out three main characteristics of big data: volume, velocity and variety. Big data help to create value for companies through the incremental improvement and optimisation of current business practices, processes and services, and through innovation in new products and business models based on data use (Hartmann et al. 2016).

Technological development has enabled significant improvements in CRM systems and big data. The following section discusses the connection between ICT and big data, with special attention to evidence in the tourism industry.

12.3 ICT and Big Data in Tourism

The tourist sector has been characterized by the development and implementation of new technologies, in particular, Information and Communication Technologies (ICT) (Buhalis 1998; Buhalis and Amaranggana 2015). ICT can cover "*all forms of technology utilized to create, capture, manipulate, communicate, exchange, present, and use information in its various forms –business data, voice conversations, still images, motion pictures, multimedia presentations, and other forms, including those not yet conceived.*" (Ryssel et al. 2004, p. 198). In recent years, there has been exponential growth in the use and implementation of these technologies in the tourism industry, and they have become an essential element in a company's competitive improvement by contributing to the efficient management of information both inside the organisation and in relations with customers. The term "Smart tourism" has been proposed to describe the increasing reliance of tourism destinations, their industries and their tourists on emerging forms of ICT which can be used to transform massive amounts of data into value propositions (Gretzel et al. 2015). All these ICT-driven changes in the tourist consumption experience provide a clear research opportunity (Neuhofer et al. 2014).

ICT tools and applications have enabled tourism firms to become 'smarter' about how to increase their performance and competitiveness by improving the efficiency of their business functions and processes (Sigala and Marinidis 2012), and by redefining their business model and the way they propose to create customer value (Gretzel et al. 2015).

From the customer perspective, ICT have enabled rapid growth of the services industry through service improvements (e.g. Lowson 2001). Recent research has focused on customer perception of the degree of a firm's ICT advancement, defined as the extent to which a company adopts the most sophisticated technology as an indicator of the company's proactive approach to providing solutions for their customers ahead of competitors (Wu et al. 2006).

In recent decades, many firms have developed CRM systems as a combination of *people*, *processes* and *technology* (Mendoza et al. 2007). The technology component of CRM locates, collects, stores and analyses data on customer patterns, interprets customer behaviour and develops predictive models, ensuring timely responses, effective customised communications and the delivery of customised products and services to individual customers (Mendoza et al. 2007). Aligning technology investment with human talent and organisational resources may be difficult (Weill and Ross 2009), and the advent of big data has added even more data, with different characteristics, for consideration in decision-making. Numerous technologies support big data creation with the goal of gaining insight from massive amounts of data, the core of smart tourism initiatives (Gretzel et al. 2015). Combining and integrating a firm's big data with frontline employees' (FLE) small data collected through CRM systems is crucial to absorbing and applying big data knowledge in frontline management (Lam et al. 2017). "Small data" refers to FLE data collected through interactions and relationships with customers. The knowledge stemming from small data needs to be complemented by the firm's big data knowledge and vice versa, when (1) customers become increasingly sophisticated, (2) FLEs have less autonomy and the organization becomes more centralized, and (3) competitive intensity is strong (Lam et al. 2017). In these contexts, which largely describe the current characteristics of the hospitality industry, FLE small data are simply not enough, and the firm and its FLEs need to actively share tacit knowledge and combine knowledge across organizational members (Lam et al. 2017) to support internal business decisions (Davenport and Dyché 2013).

Scholars have, however, paid little attention to customers' perceptions of advanced technology, whereas much of the theoretical and empirical literature in the hotel context has focused on managers' perceptions of different technology solutions. The hospitality industry has managed to adjust itself to the current digital environment (Lee et al. 2003), where consumers are adopting a more proactive attitude as they explicitly express their opinions and perceptions (Gurău 2008; Sigala 2012). Therefore, as Mulhern (2009) suggests, we believe that the correct way to understand this new digital landscape where information, marketing communications, and advanced technology converge, is from the customer perspective. Consequently, we adopt a customer-centric approach and examine our research model from the hotel guests' point of view.

12.4 Hypotheses and Proposed Model

The literature states that ICT development is positively influenced by companies' innovation decisions (Reinartz et al. 2011), because technology allows innovation processes to develop, and improves the degree to which the new product meets customers' requirements (Arvanitis et al. 2013). In the services sector, ICT has played a fundamental role in the development of innovation as a facilitator, as a support infrastructure for the smooth implementation of innovation and business management, allowing cost savings and improving the coordination of companies' internal and external activities, among others (Scupola 2014). As a result, it is claimed that innovation in services requires investment in ICT to produce new products and services, and to improve business processes (Potts and Mandeville 2007).

The volume, velocity, and variety of big data require unique analytical tools, skills, and information technology. When the firm and its FLEs are highly capable of combining and integrating big data with small data knowledge, the firm's ability to discriminate relevant information from messy data is much improved (Lam et al. 2017). In contrast, frontline employees collect small data, extract information from it, and accumulate knowledge through their direct interactions and relationships with customers (Lam et al. 2017), gathering unique and context-specific data and knowledge about customer needs, incidents in service delivery, ways to improve service quality, and customer sentiment and preferences (Santos-Vijande et al. 2016; Ye et al. 2012).

Applications of big data require the use of technologies that complement FLEs' traditional service provision (e.g., the use of wearable technologies in service encounters) or even replace FLEs (e.g., check-in robots at the Japanese Henn-na Hotel; Wong 2015). The technology component in CRM systems—as well as in big data applications—is required to collect and analyse data on customer patterns, interpret customer behaviour and develop predictive models (Mendoza et al. 2007). Therefore, most of the processes in current CRM projects and big data developments are handled with the support of ICT (Rahimi 2017) to enhance insight and decision making to create value (Gartner 2012), enabled by relational innovation in some cases.

In retailing, several studies argue that technological alignment between retailers and their providers is directly related to perceived benefits from their relations and inversely related to costs and sacrifices associated to those relations (Ryssel et al. 2004; Vize et al. 2013). Therefore, relational innovation may allow firms to create value for their customers through technology. Extrapolating this notion to the hospitality industry, we posit:

H1: Hotel relational innovation positively influences hotel guest perceptions of hotel ICT advancement.

Enhancing the brand's value, or "equity" is critical to successful hotel brand management (Bailey and Ball 2006). The main brand equity conceptualization from the consumer perspective, provided by Aaker (1991, p. 15), defines brand equity as "a set of brand assets and liabilities linked to a brand, its name and symbol that adds to or subtracts from the value provided by a product or service to a firm and/or to that firm's customers". He identifies four dimensions of brand equity, namely brand awareness, perceived quality, brand associations (i.e., image), and brand loyalty.

Innovation involves two main words: trendy and popular (Wang et al. 2008) that can create and enhance value for customers and thus contribute to the creation of brand equity (Aaker 2007). Moreover, since relational innovation may be applied to enhancing service quality with guests' information gathered through CRM solutions (Sigala and Connolly 2004) and big data (Hartmann et al. 2016), it may contribute to hotel brand differentiation. Most innovative retailers are therefore expected to enjoy superior store brand awareness and enhanced image, thus contributing to brand equity.

Although consumer-based brand equity is commonly seen as multi-dimensional within the marketing literature (e.g., Aaker 1991; Keller 1993), there is ongoing discussion as to whether the principles of branding within goods marketing can be directly applied to service dominant brands such as hotels. For instance, Aaker (1991) recognises perceived quality as one of the components of brand equity but does not specify whether this refers to goods or services, and therefore this study does not state whether the brand equity model is suitable for assessing service-dominant brand equity. Indeed, applications of goods-based brand equity models show poor validity in the tourism industry (Boo and Baloglu 2009). Goods-based branding models need adjusting to accommodate the unique characteristics of services (e.g., intangibility, variability, inseparability, heterogeneity) when evaluating service dominant brands (Blankson and Kalafatis 1999; O'Cass and Grace 2004; Kayaman and Arasli 2007). Based on the above discussion, we consider hotel brand equity as a unidimensional construct (overall brand equity) and we hypothesize that:

H2: Hotel relational innovation positively influences hotel guest perceptions of hotel overall brand equity.

Consumers nowadays are characterised by being highly connected (Wang et al. 2014, 2016; Murphy et al. 2016) continuously using the Internet (Buhalis and Law 2008), moving from the status of mere viewer of contents to become generators of information on social networks, blogs, and so on, and participants in the creation of tourist experiences. Under this approach of conjoint creation of value by the company and the consumer (Vargo and Lusch 2004), technology emerges as one of the most notable elements in the co-creation of tourist experiences by allowing consumers to become more involved in generating their experiences, thereby enhancing customer value (Binkhorst and Den Dekker 2009) and contributing to create brand equity.

The implementation of technologies in hotels delivers notable advantages in management knowledge (Li et al. 2012), competition, increased profitability, cost reduction, efficiency, and information-sharing (Lee et al. 2003; Ham et al. 2005). In addition, Gilbert and Powell-Perry (2001) report that the Web is an effective marketing relationship tool, while Lee et al. (2003) find that, according to hotel managers' opinions and beliefs, technology can also enhance the quality of service, improve the overall image of the hotel, and encourage customer loyalty.

Relational innovations, intangible in nature, have been associated with increased levels of trust, relationship quality and loyalty between customers and service providers, the three dimensions Cassivi et al. (2008) consider to measure the construct. Relational innovation may be based on CRM systems that play a vital role in increasing guest satisfaction, loyalty, and retention (Lo et al. 2010) and big data developments that aim to enhance insight and decision making to create value (Gartner 2012) thereby potentially contributing to strengthen relational ties with the hotel. Therefore, we posit:

H3: Hotel relational innovation positively influences hotel guest perceptions of the strength of relational ties.

Customer retention is a main firm objective, as it affects company profitability due to the fact that it is less expensive to maintain customers than to acquire new ones (Mendoza et al. 2007). Loyalty has been defined as the combination of a positive attitude and purchase repetition (Dick and Basu 1994). In the context of services, Salegna and Goodwin (2005, p. 54) consider loyalty as "the desire to go to the service provider as the result of a high level of satisfaction, high emotional commitment and continued repeat purchase behavior". In hotels, CRM solutions have been implemented to seek, gather and store information on guests that serves to identify and retain the most profitable customers and improve the profitability of less profitable customers (Sigala and Connolly 2004), because CRM uses technology to further the company's insight and involvement with its customers (Mendoza et al. 2007). Big data provide more, and possibly better, data for CRM decision-making, enabling relational innovation that ultimately may contribute to customer retention and loyalty.

Moreover, according to the relational perspective of innovation in marketing channels pointed out by Musso (2010), innovation can occur in relationships with end-customers through *customer care initiatives*, that is, all activities aimed at strengthening the relationship with the end user, such as the use of loyalty cards. Hotel relational innovativeness is therefore expected to influence consumer perceptions of guest loyalty. Thus we hypothesize:

H4: Hotel relational innovation positively influences customer loyalty towards the hotel.

The theory of resources and capabilities argues that ICT have the potential to generate value (Bruque et al. 2003), since they are (1) a scarce and valuable strategic resource, (2) complementary to other resources, since the value generation of ICT requires interaction with other systems and processes, and (3) difficult to imitate—not the technology solution in itself, but the interaction with staff and the organization.

The literature suggests that technology can improve some of Aaker's (1991) brand equity dimensions like brand image and perceived quality, thus becoming a source of long-lasting relationships with clients (e.g., Lee et al. 2003; Kim and Hardin 2010). There is evidence to show that technologies can enhance hotel image (Ruiz-Molina et al. 2011). The use of technologies based on big data potentially enhances how customers perceive the servicescape (Lam et al. 2017). When these big data technologies are integrated with other tangible elements such as furniture and artefacts (Bitner 1992) at the customer interface, customers will perceive the servicescape as more modern and "digital ready" (Lam et al. 2017), thus potentially contributing to guest perceptions of hotel brand equity.

Furthermore, modern technological applications help to improve hotel service quality for increasingly demanding customers (Lee et al. 2003; Law and Jogaratnam

2005; Ruiz-Molina et al. 2011). For instance, real-time identification of individual customers through beacons and similar technology can help improve service quality (Lam et al. 2017). Hotels are thus embracing the use of new technologies to improve their services and to increase customer loyalty (Lee et al. 2003). Studies in the hotel context confirm the impact of virtual world technology on word-of-mouth about the hotel (e.g., Kim and Hardin 2010).

As a number of authors have suggested, academics should also consider the role of advanced technology solutions in order to understand consumers and their relationship with the brand and the company (e.g., Schultz 1999; Reid 2002; Kliatchko 2009). Therefore we posit:

H5: Hotel guest perceptions of hotel ICT advancement positively influences hotel overall brand equity.

Technological advances enable firms to faster and more effectively collect, analyse and use customer intelligence for better understanding of the customer and providing personalised interactions through the entire customer life cycle (Küpper et al. 2014; Sigala 2016). Changes in technology and consumer behaviour are transforming CRM from a transactional to a conversational approach, known as social CRM or CRM 2.0 (Diffley and McCole 2015; Malthouse et al. 2013) empowering customers as the co-creators of relationships (Sigala 2016). Firms are no longer in control of the customer relationship; instead, individual customers and virtual communities are now driving the conversation and influencing brand image and relations (Dessart et al. 2015). Social customers expect personalised customer service and information on any device, at any place and time (Sigala 2015), and they become strong brand advocates when customer satisfaction is achieved (Sigala 2016). As technologies give customers control over relationships, several authors have recognised the need to change the scope and aim of CRM, so that, instead of managing customer relations to generate business value, management's priority and the challenges of social CRM are how to manage and motivate customer engagement to ensure that customers participate in value co-creation activities (Sigala 2016). Furthermore, personalized treatment through customer identification enabled by CRM systems may improve customer perceptions of frontline employees' empathy and responsiveness, and increase customer confidence and trust (Lam et al. 2017), "basic relationship building blocks" according to Morgan and Hunt (1994) of the supplier's relationship with its customer. Moreover, big data can be used to identify customers along with their preferences and habits and their situational and emotional state (Lam et al. 2017), thus contributing to enhance guests' relationships with the hotel brand.

The use of technology as a substitute for frontline employees (Lam et al. 2017), however, usually limits response options, which might lead to customer frustration. In addition, today's technology prevents the provision of social benefits that might be essential for some customers (e.g., Hennig-Thurau et al. 2006). Moreover, without small data, big data may result in more big costs than useful benefits (Lam et al. 2017:25). Therefore, we posit:

H6: Hotel guest perceptions of hotel ICT advancement positively influences the strength of relational ties.

Brand equity is seen as a relational market-based asset because it exists outside the firm and resides in end user relationships with brands (Falkenberg 1996; Hooley et al. 2005; Srivastava et al. 1998, 2001). The branding literature defines it as a relational asset (Aaker 1991; Keller 1993), because much of its value is a result of the brand's external relationships with other members of the value chain (e.g. distribution system and end users), and in the market place, it derives from the set of brand associations and behaviours that have been developed towards the brand.

Moreover, following Broyles et al. (2009), five consequences of brand equity have been discussed in the literature: (1) reduced anticipated risk concerning a brand purchase decision (Guerrero et al. 2000; Lassar et al. 1995), (2) increased anticipated confidence in a brand purchase decision (Aaker 1996), (3) anticipated satisfaction with the product, (4) reduced difficulty with regard to the purchase decision process (Aaker 1992; De Chernatony and Dall'Olmo 1997) and (5) purchase behaviour (Aaker 1991; Farquhar 1989; Guerrero et al. 2000; Keller 1993). These consequences are expected to reinforce the relationship between customer and service provider.

All in all, as a relational market-based asset, brand equity may be expressed as a function of brand-consumer relationships (Ambler 1997), and as such we understand that it may be positively related to the strength of relational ties. Consequently, we hypothesize that:

H7: Hotel guest perceptions of hotel overall brand equity positively influences the strength of relational ties.

Firms commit considerable resources to relationship marketing strategies with the aim of increasing relationship satisfaction and loyalty (Balaji 2015). Such investments create expectations of reciprocity through financial, social, and structural ties and motivate the parties to strengthen their relationships (Hsieh et al. 2005; Balaji 2015). In the context of firm–customer relationships, customers feel compelled to respond to relationship investment strategies by increasing their loyalty towards the firm (De Wulf et al. 2001). Indeed, it has been argued that high levels of relationship investments lead to greater relationship quality and loyalty (De Wulf et al. 2001) and increased retention (Ahmad and Buttle 2001).

Several authors operationalize the overall strength of firm–customer relationships as relationship quality, comprising three components, trust, satisfaction, and commitment (Garbarino and Johnson 1999; Balaji 2015). As the number of relational ties grows, interactions between the firm and the customer increase, the firm acquires critical customer information, enabling it to respond to market changes (Jap and Ganesan 2000), identify opportunities for better meeting customer needs and building strong relationships (Gassenheimer et al. 1995). As the number of satisfactory firm–customer interactions increase, so do trust, commitment to the relationship and the duration of that relationship (Morgan and Hunt 1994).

In support of the influence of the strength of relational ties on customer loyalty, in a study conducted in an apparel retailing setting, De Cannière et al. (2010) report that buying intentions have a greater impact on purchase behaviour among customers with a strong relationship with the service provider. Generalising this finding to the hospitality industry, we hypothesize:

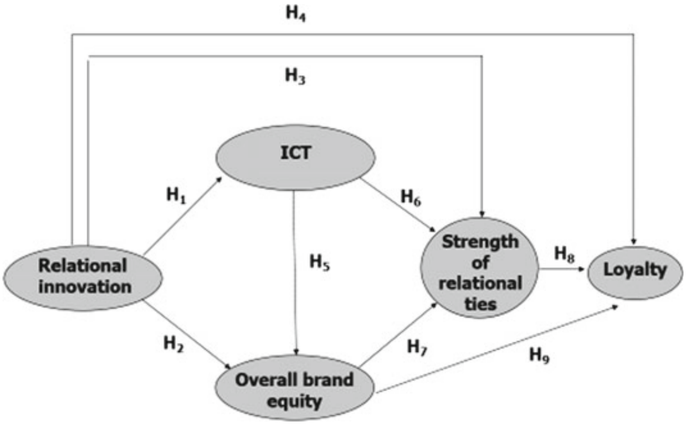


Fig. 12.1 Proposed model

H8: The strength of relational ties positively influences customer loyalty towards the hotel.

A brand enjoys positive customer-based brand equity when customers react more favourably to a product identified with a brand and the way that it is marketed than when it is not (Keller 1998). Several studies find a positive relation between overall brand equity and behavioural intentions (Buil et al. 2013; Wong 2013). In particular, a consumer with a positive perception of the brand shows greater willingness to buy the product (Luijten and Reijnders 2009; Fleck et al. 2012; Tantiseneepong et al. 2012; Thwaites et al. 2012; Chan et al. 2013; Sasmita and Mohd Suki 2015), recommend to others and pay more for the brand in comparison to other alternatives (Choi et al. 2009; Han and Kim 2010; Han et al. 2011).

In a similar vein, Keller (1998) argues that one of the characteristics of brands possessing strong brand equity is stronger brand loyalty. Traditionally, loyalty has been considered as a dimension of brand equity (Aaker 1991, 1996; Yoo et al. 2000; Pappu et al. 2005); however, according to recent empirical evidence (e.g. DeCarlo et al. 2007; Martenson 2007; Chebat et al. 2009; Fleck and Nabec 2010) customer loyalty is the ultimate consequence of brand equity. Based on the evidence for the relationship between brand equity and loyalty, we posit:

H9: Hotel overall brand equity positively influences customer loyalty towards the hotel.

Figure 12.1 summarizes graphically the above relations between the constructs identified as relevant for the present research.

12.5 Method

To empirically test the research model and hypotheses, we conducted quantitative research using survey methodology. We developed a survey instrument based on an extensive literature review. Most of our measures were adapted from previous studies to fit the current context of the research. Thus, items to measure Relational innovation were adapted from Homburg et al. (2002); the scale to measure ICT advancement was adapted from Wu et al. (2006); items to assess Overall brand equity were adapted from Yoo and Donthu (2001) and So and King (2010). Items for measuring Strength of relational ties were adapted from Oke and Idiagbon-Oke (2010), and the Loyalty scale was adapted from Kim and Kim (2005). All items were measured on a 7-point Likert scale. Classification variables to describe hotel guest profile were included at the end of the questionnaire.

The field work was conducted in April and May 2016 in the Valencia region, one of the main holiday destinations in Spain. We chose April and May for the survey because, in a short period of time, we would be able to gather data on travellers staying at the hotel during the low, medium and high seasons (Easter holidays and 1st and 2nd May bank holidays in several Spanish regions). Hotel managers were also more likely to allow researchers to interview their guests in the lobby area than during the summer period, when hotel facilities are rather busy. The study population is guests of hotels in two of the region's main cities (Valencia and Gandía). According to data from the Spanish Institute of Statistics (INE), the Valencia region received 6.5 million tourists in 2015 (4.3% more than in 2014), representing 9.5% of the total number of foreign tourists visiting Spain. Valencia is among the top 10 Spanish cities in terms of visitor numbers. Around two million tourists visited Valencia in 2016, which is an increase of 2.2% over 2015. The second most popular holiday destination in the Valencia region is Gandía. A census of hotels was elaborated from secondary information in the Official Guide of Hotels in Spain (Turespaña 2017) and Valencia Tourism Agency's directory of hotels (Agència Valenciana de Turisme 2017). This information was completed with the SABI (System of Analysis of Iberian Balance Sheets) and DUNS100.000 databases.

After selecting the hotels, permission was requested to interview hotel guests face-to-face at the hotel, in the reception or lobby area. For a ±5% sample error, at least 400 valid questionnaires were required, so interviewers approached 1175 potential respondents to finally obtain 401 valid face-to-face questionnaires from guests of 3, 4 and 5-star hotels (response rate: 34%). The number of surveys performed in Valencia and Gandía were 332 and 70 (82.59 and 17.41%, respectively). The number of surveys performed in each hotel category was proportional to the number of hotels in this category authorizing the research team to interview their guests. In particular, 154 (38.3%), 232 (57.7%) and 16 (4%) guests were interviewed in 3, 4 and 5 star-hotels. Table 12.1 shows the sample distribution according to the classification variables included in the survey.

With the collected data, first, exploratory factor analyses were conducted through principal component analysis (PCA) with VARIMAX rotation to confirm the uni-

Table 12.1 Sample profile

Classification variables	N	%
Gender		
Male	186	46.4
Female	215	53.6
Age		
18–25	43	10.7
26–35	76	19.0
36–45	116	28.9
46–55	78	19.5
56–65	46	11.5
>65	34	8.5
NA	8	2.0
Educational level		
No studies/primary school	62	15.5
High school	61	15.2
Vocational training	74	18.5
University	185	46.1
Master/Ph.D.	9	2.2
Other	3	0.7
NA	7	1.7
Trip motivation		
Leisure	351	87.5
Business	33	8.2
Others	9	2.2
NA	8	2.0

dimensionality of the scales used to measure the constructs. Then the partial least squares (PLS) method was used to test the measurement model and the structural equations model (SEM) with SmartPLS3 (Ringle et al. 2014). PLS-SEM is a suitable statistical technique for our research, since it assumes minimal restrictions on construct distributions (Chin et al. 2003). Furthermore, PLS-SEM is preferred over covariance-based SEM (CB-SEM) when analysing predictive research models that are in the stages of theory development (Gimbert et al. 2010; Rezaei 2015).

Table 12.2 shows the confirmation of measurement model reliability and validity. Composite reliability (CR) values are greater than the recommended cut-off value of 0.80, and average variance extracted (AVE) values are all greater than the threshold value of 0.5.

Regarding factor loadings, two items on the loyalty scale, items L1 ("I regularly visit this hotel") and L3 ("I usually use this hotel as my first choice compared to other hotels"), were eliminated because loadings were under 0.6. After purging the

Table 12.2 Confirmatory factor analysis results

Construct	Item	St. loading factor (st. error)	t	Cronbach's alpha	Composite reliability	AVE
F1. Relational innovation	RI1. This hotel adopts more innovations or new services ideas in their relations with customers than other hotels	0.952 (0.008)	120.79*	0.952	0.969	0.913
	RI2. This hotel adopts innovations or new services ideas in their relations with customers more quickly than other hotels	0.973 (0.004)	236.30*			
	RI3. This hotel adopts innovations or new services ideas in their relations with customers over time relative to other hotels	0.942 (0.011)	84.375*			
F2. ICT	ICT1. This hotel invests in technology	0.900 (0.011)	78.707*	0.918	0.942	0.803
	ICT2. This hotel has the most advanced technology	0.914 (0.014)	65.827*			
	ICT3. Relative to its competitors, the technology of this hotel is more advanced	0.927 (0.010)	92.471*			
	ICT4. This hotel considers my opinion as a customer on decisions involving IT coordination and development	0.842 (0.027)	31.152*			
F3. Overall brand equity	OBE1. It makes sense to stay at this hotel instead of any other hotel, even if they are the same	0.900 (0.016)	57.241*	0.953	0.966	0.878

(continued)

Table 12.2 (continued)

Construct	Item	St. loading factor (st. error)	t	Cronbach's alpha	Composite reliability	AVE
	OBE2. Even if another hotel has the same features as this one, I would prefer stay at this hotel	0.939 (0.010)	91.359*			
	OBE3. If there is another hotel as good as this hotel, I prefer to stay at this hotel	0.962 (0.005)	176.47*			
	OBE4. If another brand is not different from this hotel in any way, it seems smarter to stay at this hotel	0.946 (0.009)	106.85*			
F4. Strength of relational ties	SRT1. This hotel work towards attaining similar goals to mines: it innovates to align its objectives to mines	0.811 (0.031)	25.920*	0.936	0.950	0.760
	SRT2. I would be interested in continuing my relationship with this hotel in the future	0.824 (0.022)	36.698*			
	SRT3. Even if I had other options, I would remain in this hotel because of its ability to innovate	0.911 (0.010)	88.052*			
	SRT4. There are few problems with this hotel: it innovates to reduce or eliminate frictions with customers	0.919 (0.010)	87.531*			
	SRT5. The hotel innovates to make relationships with their customers close and personal	0.910 (0.012)	77.139*			

(continued)

Table 12.2 (continued)

Construct	Item	St. loading factor (st. error)	t	Cronbach's alpha	Composite reliability	AVE
	SRT6. Thanks to the innovations of this hotel, there is a good relationship between me and this hotel	0.850 (0.021)	40.052*			
F5. Loyalty	L2. I intend to visit this hotel again	0.789 (0.027)	29.475*	0.841	0.893	0.677
	L4. I am satisfied with the visit to this hotel	0.819 (0.033)	24.939*			
	L5. I would recommend this hotel to others	0.868 (0.016)	54.488*			
	L6. I would not switch to another hotel the next time	0.813 (0.019)	42.856*			

*Statistically significant at $p < 0.05$

Table 12.3 Square root of AVE and correlations between constructs

	F1	F2	F3	F4	F5
F1. Relational innovation	**0.956**				
F2. ICT	0.800	**0.896**			
F3. Overall brand equity	0.239	0.285	**0.937**		
F4. Strength of relational ties	0.656	0.576	0.318	**0.872**	
F5. Loyalty	0.315	0.338	0.794	0.394	**0.823**

Diagonal values in bold are square roots of AVE and others (off-diagonal) are correlations between variables

loyalty scale, Table 12.2 shows that all factor loadings for the constructs are higher than the recommended value of 0.7 (Hair et al. 2014). Therefore, factor loadings, AVE and CR for the reflective constructs in the final measurement model indicate sufficient convergent validity.

Discriminant validity was confirmed as each square root of AVE (on the diagonal in Table 12.3) was higher than the correlations with other constructs, according to the Fornell and Larcker (1981) criterion. Thus, our measurement for the reflective constructs is valid.

The psychometric properties of the measurement model seem to be appropriate and therefore, the structural equations model can be estimated.

12.6 Results

After ensuring that reliability, convergent validity, and discriminant validity requirements are all satisfied, the structural equations model was estimated. This stage involved determining the relationship between Relational innovation, ICT advancement, Overall brand equity, Strength of relational ties and Loyalty. The bootstrapping resampling procedure in PLS was used to estimate the structural model and test the hypotheses.

As shown in Table 12.4, Relational innovation exerts a positive and significant influence on hotel ICT advancement which in turn, positively influences overall brand equity. The latter shows a positive and significant influence on strength of relational ties and loyalty. Strength of relational ties is positively influenced by both relational innovation and overall brand equity. Loyalty is determined by overall brand equity and strength of relational ties. Thus, overall brand equity shows both a direct and an indirect effect on loyalty, mediated by the strength of relational ties. Finally, technology does not seem to directly influence the strength of relational ties, but it enhances overall brand equity which in turn, contributes positively to increasing the strength of relational ties.

These findings, however, maybe be influenced by some factors related to guest characteristics (e.g. guest age, that may affect consumer perceptions of hotels innovativeness and ICT) and hotel features (e.g. hotel category, that may be related to investment in ICT). To refine the results by guest age and hotel classification, a PLS Multigroup Analysis was performed for three age groups (i.e. under 36, between

Table 12.4 Structural equation model results and hypothesis testing

Structural relation	Hypothesis	Standardized β	Bootstrapping t
H_1: Relational innovation → ICT	Supported	0.800	33.53[a]
H_2: Relational innovation → Overall brand equity	Not supported	0.032	0.40
H_3: Relational innovation → Strength of relational ties	Supported	0.538	7.41[a]
H_4: Relational innovation → Loyalty	Not supported	0.051	1.11
H_5: ICT → Overall brand equity	Supported	0.258	2.93[a]
H_6: ICT → Strength of relational ties	Not supported	0.100	1.16
H_7: Overall brand equity → Strength of relational ties	Supported	0.162	3.13[a]
H_8: Strength of relational ties → Loyalty	Supported	0.121	2.34[b]
H_9: Overall brand equity → Loyalty	Supported	0.741	28.93[a]

ICT: $R^2 = 0.640$. $Q^2 = 0.482$; Overall brand equity: $R^2 = 0.081$. $Q^2 = 0.065$
Strength of relational ties: $R^2 = 0.462$. $Q^2 = 0.326$; Loyalty: $R^2 = 0.654$. $Q^2 = 0.413$
[a, b, c]Statistically significant at $p < 0.01$, 0.05 and 0.10, respectively

36 and 45, and over 45) and two hotel categories (i.e. three-star hotels and upscale hotels—namely 4 and 5-star hotels). The groups were defined to guarantee at least 100 respondents in each group.

Before performing the PLS-MGA, measurement invariance was addressed to determine if the measurement models specify measures of the same attribute under different conditions. To test for measurement invariance in PLS-SEM, the MICOM procedure was executed (Henseler et al. 2016). This procedure requires three steps to test for invariance of configuration and composition, as well as equality of composite mean values and variances (Henseler et al. 2016).

After establishing invariance, for the three age groups and for the two hotel category groups, the focus turned to determining whether the path coefficients of the theoretical models for the different groups are significantly different. Results for PLS-MGA by guest age and hotel category are shown in Tables 12.5 and 6, respectively.

Regarding the differences between path coefficients across guest age (Table 12.5), the impact of relational innovation on ICT is significantly lower for hotel guests older than 45 in comparison to the other age groups of younger consumers. Moreover, the relationship between relational innovation and loyalty is stronger for consumers under 36 in comparison to hotel guests over 45. This may be explained by the peculiarities of each age cohort. The age group of hotel guests under 36 largely corresponds to Generation Y (also called Millennials) and Z (Centennials), whereas hotel guests aged between 36 and 45 could be considered members of Generation X, and those over 45 may be mainly Baby boomers (Pendergast 2010; Williams et al. 2010; Strutton et al. 2011). Both Millennials and Centennials are considered digital natives (Prensky 2001) and this may explain the stronger link between relational innovation and ICT in comparison to other age cohorts.

As far as the differences between path coefficients across hotel category are concerned (Table 12.6), the impact of relational innovation on ICT is significantly higher for upscale establishments (i.e. 4 and 5 star hotels) in comparison to 3-star hotels. Similarly, the impact of overall brand equity on the strength of relational ties is stronger for upscale hotels in comparison to lower category hotels. These findings may be explained by the more intense efforts of 4 and 5-star hotels in innovation and branding in comparison to 3-star hotels. In contrast, the impact of ICT on the strength of relational ties is higher for guests of 3-star hotels in comparison to those of higher categories. This result may be explained by the fact that 3-star hotels face more intense competition in comparison to upscale hotels, and therefore may use CRM systems and other technological developments in order to strengthen relational ties with customers and retain guests.

Table 12.5 PLS-MGA examining differences across age groups

Paths	Age groups			Path coefficient differences			p-values		
	<36 n = 119	36–45 n = 116	>45 n = 158	1 versus 2	1 versus 3	2 versus 3	1 versus 2	1 versus 3	2 versus 3
H_1: Relational innovation → ICT	**0.856**	**0.797**	**0.683**	0.059	0.173	0.114	0.132	**0.001**	**0.048**
H_2: Relational innovation → Overall brand equity	0.032	0.027	0.078	0.005	0.047	0.052	0.494	0.594	0.614
H_3: Relational innovation → Strength of relational ties	**0.529**	**0.514**	**0.509**	0.015	0.020	0.005	0.465	0.464	0.482
H_4: Relational innovation → Loyalty	**0.171**	−0.008	−0.056	0.179	0.227	0.048	0.072	**0.019**	0.323
H_5: ICT → Overall brand equity	0.256	**0.407**	0.193	0.151	0.063	0.214	0.752	0.381	0.098
H_6: ICT → Strength of relational ties	0.054	**0.199**	0.109	0.145	0.055	0.089	0.789	0.598	0.317
H_7: Overall brand equity → Strength of relational ties	**0.179**	**0.227**	0.108	0.049	0.071	0.119	0.660	0.270	0.134
H_8: Strength of relational ties → Loyalty	0.030	0.144	**0.198**	0.114	0.167	0.053	0.826	0.941	0.689
H_9: Overall brand equity → Loyalty	**0.720**	**0.773**	**0.763**	0.054	0.044	0.010	0.806	0.784	0.419

1: younger than 36, 2: between 36 and 45, 3: older than 45
Bold values are significant at $p < 0.05$

Table 12.6 PLS-MGA examining differences across hotel category

Paths	3 stars n = 154	4/5 stars n = 248	Path coefficient diff.	*p*-values
H_1: Relational innovation → ICT	**0.635**	**0.831**	0.195	**0.999**
H_2: Relational innovation → Overall brand equity	−0.012	0.105	0.116	0.788
H_3: Relational innovation → Strength of relational ties	**0.431**	**0.602**	0.171	0.900
H_4: Relational innovation → Loyalty	0.058	0.067	0.010	0.541
H_5: ICT → Overall brand equity	**0.345**	0.169	0.176	0.122
H_6: ICT → Strength of relational ties	**0.279**	−0.002	0.281	**0.033**
H_7: Overall brand equity → Strength of relational ties	−0.006	**0.241**	0.247	**0.996**
H_8: Strength of relational ties → Loyalty	0.111	0.127	0.016	0.562
H_9: Overall brand equity → Loyalty	**0.775**	**0.717**	0.058	0.106

Bold values are significant at $p < 0.05$

12.7 Conclusions

The aim of the present research was to contribute to the tourism marketing literature by analysing the connections between relational innovation, hotel technological advancement, brand equity, strength of relational ties and guest loyalty towards the hotel from the consumer perspective.

The results of our quantitative research based on a sample of guests at Spanish hotels highlight the relevance of several constructs, traditionally neglected in the literature, on brand equity creation and customer loyalty. Firstly, relational innovation emerges as a determinant of the strength of relational ties between hotel and guest and, through the influence of those ties on technological advancement, of store brand equity. Secondly, the strength of relational ties makes a remarkable contribution to guest loyalty towards the hotel. Thirdly, hotel level of technological advancement contributes to create brand equity. And lastly, regarding the discussion in the literature about the relation between brand equity and customer loyalty, the present research supports the notion that loyalty is an outcome of brand equity rather than one of its dimensions.

These findings have both theoretical and managerial implications. The present research contributes to the literature by testing a model to examine the connections between relational innovation, technological advancement, brand equity, strength of relational ties and guest loyalty. To the best of our knowledge these aspects have not been analysed simultaneously in the context of the tourism industry. From a managerial standpoint, the present study provides evidence for the influence of relational innovation and technology in generating hotel brand equity and strengthening the links of the hotel guest with the establishment. We consider therefore that, in view of the ultimate positive impact on customer loyalty towards the establishment, hotels should invest in CRM systems and big data developments to innovate in the relationship with their guests and to design and develop new customized services to differentiate themselves from their competitors, thereby enhancing hotel brand equity and strengthening relational ties with their customers. Moreover, differences in some of the hypothesized relations across age groups and hotel category suggest the need for hotel managers to prioritize their investments depending on the most typical profile of hotel guests and hotel positioning, closely related to hotel category.

The present research, however, is not without its limitations and they, in turn, suggest further research lines. Firstly, technology has been measured as guest perceptions of hotel technological advancement. This measurement needs further refinement and future research should analyse guest perceptions of certain aspects of technology and their impact on customer variables. Indeed, researchers have called for further study on customer variables relevant to big data marketing (Lam et al. 2017). We consider that further research should address customer acceptance of big data marketing (information sharing, providing useful inputs in co-creating products and services), customer perceptions of hotel CRM systems, and customer adoption of new technology and services, among others. Even though we attempted to gather these aspects in guest perception of hotel relational innovation, more refined measurement instruments need to be developed to assess guest perceptions of big data and CRM use by the hotel.

Moreover, while research is already being conducted on how technology enhances tourism experiences, little attention has been paid to the potential drawbacks of intensive use of ICT solutions (Gretzel et al. 2015). Costs and sacrifices of technology use should be introduced in the model to analyse their role in the relations between relational innovation, strength of relational ties, overall brand equity and loyalty.

Lastly, a larger sample would have enabled further refinement of results by trip motivation and guest cohort, distinguishing Millennials from Centennials and other age cohorts, since the literature highlights the need to urgently update the design and structure of hotel CRM strategies to define value propositions appealing to Millennials (Bowen and McCain 2015) who seek more customized experiences, flexibility and variety (Sigala 2011); and are always connected for sharing, communicating and engaging with peers and social networks using various technological devices.

Acknowledgements The authors are very grateful for the support of Spanish Ministry of Education and Science projects ECO2013-43353-R and ECO2016-76553-R.

References

Aaker DA (1991) Managing brand equity. Free Press, New York
Aaker DA (1992) The value of brand equity. J Bus Strategy 13(4):27–32
Aaker DA (1996) Building strong brand. Free Press, New York
Aaker D (2007) Innovation: brand it or lose it. Calif Manag Rev 50(1):8–24
Agència Valenciana de Turisme (2017) *Hoteles*. Available at http://www.turisme.gva.es/opencms/opencms/turisme/es/contents/hoteles/BuscaHoteles.jsp. Accessed 22 Jan 2018
Ahmad R, Buttle F (2001) Customer retention: a potentially potent marketing management strategy. J Strateg Mark 9(1):29–45
Aiden E, Michel J-B (2014) The predictive power of big data. Newsweek. Available at http://www.newsweek.com/predictive-power-big-data-225125. Accessed 22 Jan 2018
Amabile TM (1983) The social psychology of creativity. Springer, New York
Ambler T (1997) How much of brand equity is explained by trust? Manag Decis 35(4):283–292
Anttiroiko AV, Valkama P, Bailey SJ (2014) Smart cities in the new service economy: building platforms for smart services. AI Soc 29(3):323–334
Arvanitis S, Loukis E, Diamantopoulou V (2013) The effect of soft ICT capital on innovation performance of Greek firms. J Enterp Inf Manag 26(6):679–701
Bailey R, Ball S (2006) An exploration of the meanings of hotel brand equity. Serv Ind J 26(1):15–38
Balaji MS (2015) Investing in customer loyalty: the moderating role of relational characteristics. Serv Bus 9(1):17–40
Binkhorst E, Den Dekker T (2009) Agenda for co-creation tourism experience research. J Hospitality Mark Manag 18(2–3):311–327
Bitner MJ (1992) Servicescapes: the impact of physical surroundings on customers and employees. J Mark 56(April):57–71
Blankson C, Kalafatis SP (1999) Issues and challenges in the positioning of service brands: a review. J of Product Brand Manag 8(2):106–118
Boo S, Busser J, Baloglu S (2009) A model of customer-based brand equity and its application to multiple destinations. Tour Manag 30(2):219–231
Bowen JT, McCain SLC (2015) Transitioning loyalty programs. Int J Contemp Hospitality Manag 27(3):415–430
Brochado A, Rita P, Margarido A, Margarido A (2016) High tech meets high touch in upscale hotels. J Hospitality Tourism Technol 7(4):347–365
Broyles SA, Schumann DW, Leingpibul T (2009) Examining brand equity antecedent/consequence relationships. J Mark Theor Prac 17(2):145–161
Bruque S, Moyano J, Vargas A, Hernández MJ (2003) Ownership structure, technological endowment and competitive advantage: do democracy and business fit? Technol Anal Strateg Manag 15(1):65–79
Buhalis D (1998) Strategic use of information technologies in the tourism industry. Tour Manag 19(5):409–421
Buhalis D, Amaranggana A (2015) Smart tourism destinations enhancing tourism experience through personalisation of services. In: Tussyadiah I, Inversini A (eds) Information and communication technologies in tourism 2015. Springer, Cham, pp 377–389
Buhalis D, Law R (2008) Progress in information technology and tourism management: 20 years on and 10 years after the Internet—the state of eTourism research. Tour Manag 29(4):609–623
Buil I, Martínez E, de Chernatony L (2013) The influence of brand equity on consumer responses. J Consum Mark 30(1):62–74
Cassivi L, Hadaya P, Lefebvre E, Lefebvre LA (2008) The role of collaboration on process, relational, and product innovations in a supply chain. Int J e-Collaboration 4(4):11–32
Chan K, Leung Ng Y, Luk EK (2013) Impact of celebrity endorsement in advertising on brand image among Chinese adolescents. Young Consum 14(2):167–179
Chebat J, El Hedhli K, Sirgy MJ (2009) How does shopper-based mall equity generate mall loyalty? A conceptual model and empirical evidence. J Retail Consum Serv 16(1):50–60

Chin WW, Marcolin BL, Newsted PR (2003) A partial least squares latent variable modeling approach for measuring interaction effects: results from a Monte Carlo simulation study and an electronic mail emotion/adoption study. Inf Syst Res 14(2):189–217
Choi G, Parsa HG, Sigala M, Putrevu S (2009) Consumers' environmental concerns and behaviors in the lodging industry: a comparison between Greece and the United States. J Qual Assur Hospitality Tourism 10(2):93–112
Dantuma LMY, Hawkins RW (2001) E-commerce in the logistics sector: assessing the effects on the logistics value chain. TNO and Telematica Instituut, Delft
Davenport TH, Dyché J (2013) Big data in big companies. Int Inst Anal 3:1–31
De Cannière MH, De Pelsmacker P, Geuens M (2010) Relationship quality and purchase intention and behavior: the moderating impact of relationship strength. J Bus Psychol 25(1):87–98
De Chernatony L, Dall'Olmo R (1997) Brand consultants' perspectives on the concept of 'the brand'. Mark Res Today 25:45–53
De Wulf K, Odekerken-Schröder G, Iacobucci D (2001) Investments in consumer relationships: a crosscountry and cross-industry exploration. J Mark 64(4):33–50
DeCarlo TE, Laczniak RN, Motley CM, Ramaswami S (2007) Influence of image and familiarity on consumer response to negative word-of-mouth communication about retail entities. J Mark Theor Pract 15(1):41–51
Den Hertog P (2000) Knowledge-intensive business services as co-producers of innovation. Int J Innov Manag 4(4):491–528
Dessart L, Veloutsou C, Morgan Thomas A (2015) Consumer engagement in online brand communities. J Product Brand Manag 24(1):28–42
Dick AS, Basu K (1994) Customer loyalty: toward an integrated conceptual framework. J Acad Mark Sci 22(2):99–113
Diffley S, McCole P (2015) Extending customer relationship management into a social context. Serv Ind J 35(11–12):591–610
Djellal F, Gallouj F (2010) Services, innovation and performance: general presentation. J Innov Econ Manag 1:5–15
Drejer I (2004) Identifying innovation in surveys of services: a Schumpeterian perspective. Res Policy 33(3):551–562
Dupuis M (2002) Innovation dans la distribution. Revue Française Du Marketing 188:61–68
Falkenberg AW (1996) Marketing and the wealth of firms. J Macromarketing 16(1):4–24
Farquhar PH (1989) Managing brand equity. Mark Res 1(3):24–33
Fleck N, Korchia M, Le Roy I (2012) Celebrities in advertising: looking for congruence or likability? Psychol Mark 29(9):651–662
Fleck N, Nabec L (2010) L'enseigne: un capital pour le distributeur. Manag Avenir 38(8):14–32
Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. J Mark Res XVIII:39–50
Ganesan S, George M, Jap S, Palmatier RW, Weitz B (2009) Supply chain management and retailer performance: emerging trends, issues, and implications for research and practice. J Retail 85(1):84–94
Garbarino E, Johnson MS (1999) The different roles of satisfaction, trust, and commitment in customer relationships. J Mark 63:70–87
Gartner (2012) IT glossary—big data. Available at www.gartner.com/it-glossary/big-data/. Accessed 22 Jan 2018
Gassenheimer JB, Calantone RJ, Scully JI (1995) Supplier involvement and dealer satisfaction: implications for enhancing channel relationships. J of Bus Ind Mark 10:7–19
Gilbert D, Powell-Perry J (2001) Exploring developments in web based relation-ship marketing within the hotel industry. J Hospitality Mark Manag 9(3):141–159
Gimbert X, Bisbe J, Mendoza X (2010) The role of performance measurement systems in strategy formulation processes. Long Range Plan 43(4):477–497
Gretzel U, Sigala M, Xiang Z, Koo C (2015) Smart tourism: foundations and developments. Electr Markets 25(3):179–188

Grissemann U, Plank A, Brunner-Sperdin A (2013) Enhancing business performance of hotels: the role of innovation and customer orientation. Int J Hospitality Manag 33:347–356
Guerrero L, Colomer Y, Guardia MD, Xicola J, Clotet R (2000) Consumer attitude towards store brands. Food Qual Prefer 11(5):387–395
Gurău C (2008) Integrated online marketing communication: implementation and management. J Commun Manag 12(2):169–184
Guttentag DA (2010) Virtual reality: applications and implications for tourism. Tour Manag 31(5):637–651
Hair JF, Hult GTM, Ringle C, Sarstedt M (2014) A primer on partial least squares structural equation modeling (PLS-SEM). SAGE Publication Inc, California
Ham S, Kim WG, Jeong S (2005) Effect of information technology on performance in upscale hotels. Int J Hospitality Manag 24(2):281–294
Han H, Hsu LTJ, Lee JS, Sheu C (2011) Are lodging customers ready to go green? An examination of attitudes, demographics, and eco-friendly intentions. Int J Hospitality Manag 30(2):345–355
Han H, Kim Y (2010) An investigation of green hotel customers' decision formation: developing an extended model of the theory of planned behavior. Int J Hospitality Manag 29(4):659–668
Hartmann PM, Zaki M, Feldmann N, Neely A (2016) Capturing value from big data—a taxonomy of data-driven business models used by start-up firms. Int J Oper Prod Manag 36(10):1382–1406
He W, Tian X, Chen Y, Chong D (2016) Actionable social media competitive analytics for understanding customer experiences. J Comput Inf Syst 56(2):145–155
Hennig-Thurau T, Groth M, Paul M, Gremler DD (2006) Are all smiles created equal? How emotional contagion and emotional labor affect service relationships. J Mark 70:58–73
Henseler J, Ringle CM, Sarstedt M (2016) Testing measurement invariance of composites using partial least squares. Int Mark Rev 33(3):405–431
Homburg C, Hoyer WD, Fassnacht M (2002) Service orientation of a retailer's business strategy: dimensions, antecedents, and performance outcomes. J Mark 66(4):86–101
Hooley GJ, Greenley GE, Cadogan JW, Fahy J (2005) The performance impact of marketing resources. J Bus Res 58(1):18–27
Hsieh Y, Chiu H, Chiang M (2005) Maintaining a committed online customer: a study across search experience-credence products. J Retail 81(1):75–82
Hu Y, Zhao S (2010) A case study of online retail innovation system on Alibaba Taobao. In: IEEE computer society, 2010 international conference on E-Business and E-Governmen. IEEE Computer Society, Washington, DC, pp 224–227
Jacob M, Tintoré J, Aguiló E, Bravo A, Mulet J (2003) Innovation in the tourism sector: results from a pilot study in the Balearic Islands. Tourism Econ 9(3):279–295
Jap SD, Ganesan S (2000) Control mechanisms and the relationship life cycle: implications for safeguarding specific investments and developing commitment. J Mark Res 37(2):227–245
Kayaman R, Arasli H (2007) Customer based brand equity: evidence from the hotel industry. Managing Serv Q 17(1):92–109
Keller KL (1993) Conceptualizing, measuring, and managing customer-based brand equity. J Mark 57:1–22
Keller KL (1998) Strategic brand management. Prentice-Hall, Upper Saddle River
Kim J, Hardin A (2010) The impact of virtual worlds on word-of-mouth: improving social networking and servicescape in the hospitality industry. J Hospitality Mark Manag 19(7):735–775
Kim HB, Kim WG (2005) The relationship between brand equity and firms' performance in luxury hotels and restaurants. Tour Manag 26(4):549–560
Kliatchko J (2009) IMC 20 years after: a second look at IMC definitions. Int J Integr Mark Commun 1(2):7–12
Krizaj D, Brodnik A, Bukovec B (2014) A tool for measurement of innovation newness and adoption in tourism firms. Int J Tourism Res 16(2):113–125
Küpper T, Lehmkuhl T, Jung R, Wieneke R (2014) Features for social CRM technology. In: 20th Americas conference on information systems. Savannah, USA, pp 1–10

Lam SK, Sleep S, Hennig-Thurau T, Sridhar S, Saboo AR (2017) Leveraging frontline employees' small data and firm-level big data in frontline management: an absorptive capacity perspective. J Serv Res 20(1):12–28

Laney D (2001) 3-D data management: controlling data volume, velocity and variety. Gartner report [blog]. Available at http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. Accessed 22 Jan 2018

Lassar W, Mittal B, Sharma A (1995) Measuring customer-based brand equity. J Consum Mark 12(4):11–19

Law R, Jogaratnam G (2005) A study of hotel information technology applications. Int J Contemp Hospitality Manag 17(2):170–180

Lee SC, Barker S, Kandampully J (2003) Technology, service quality, and customer loyalty in hotels: Australian managerial perspectives. Managing Serv Q 13(5):423–432

Li L, Lee HA, Law R (2012) Technology-mediated management learning in hospitality organizations. Int J Hospitality Manag 31(2):451–457

Lin C (2015) Conceptualizing and measuring consumer perceptions of retailer innovativeness in Taiwan. J Retail Consum Serv 24:33–41

Lo AS, Stalcup LD, Lee A (2010) Customer relationship management for hotels in Hong Kong. Int J Contemp Hospitality Manag 22(2):139–159

Lowson RH (2001) Retail operational strategies in complex supply chains. Int J Logistics Manag 12(1):97–111

Luijten T, Reijnders W (2009) The development of store brands and the store as a brand in supermarkets in the Netherlands. Int Rev Retail Distrib Consum Res 19(1):45–58

Malthouse EC, Haenlein M, Skiera B, Wege E, Zhang M (2013) Managing customer relationships in the social media era. J Interact Mark 27:270–280

Martenson R (2007) Corporate brand image, satisfaction and store loyalty. A study of the store as a brand, store brands and manufacturer brands. Int J Retail Distrib Manag 35(7):544–555

Martínez-Ros E, Orfila-Sintes F (2009) Innovation activity in the hotel industry. Technovation 29(9):632–641

Mattsson J, Orfila-Sintes F (2014) Hotel innovation and its effect on business performance. Int J Tourism Res 16(4):388–398

Mauri C (2003) Card loyalty. A new emerging issue in grocery retailing. J Retail Consum Serv 10(1):13–25

Mendoza L, Marius A, Pérez M, Grimán A (2007) Critical success factors for a customer relationship management strategy. Inf Softw Technol 49:913–945

Morabito V (2015) Big data and analytics. Springer International Publishing, Berlin

Morgan RM, Hunt SD (1994) The commitment–trust theory of relationship marketing. J Mark 58:20–38

Mulhern F (2009) Integrated marketing communications: from media channels to digital connectivity. J Mark Commun 15(2/3):85–101

Murphy HC, Chen MM, Cossutta M (2016) An investigation of multiple devices and information sources used in the hotel booking process. Tour Manag 52:44–51

Musso F (2010) Innovation in marketing channels. Emerg Issues Manag 1. Available at http://www.unimib.it/upload/gestioneFiles/Symphonya/lasteng/f20101/mussoeng12010.pdf. Accessed 22 Jan 2018

Neuhofer B, Buhalis D (2012) Understanding and managing technology-enabled enhanced tourist experiences. In: The 2nd advances in hospitality and tourism marketing & management Conference Proceedings, Corfu, Greece, 31st May–3rd June 2012

Neuhofer B, Buhalis D, Ladkin A (2014) A typology of technology-enhanced tourism experiences. Int J Tourism Res 16(4):340–350

Nicolau L, Santa-María MJ (2013) The effect of innovation on the hotel market value. Int J Hospitality Manag 32:71–79

Novelli M, Schmitz B, Spencer T (2006) Networks, clusters and innovation in tourism: a UK experience. Tour Manag 27(6):1141–1152

O'Cass A, Grace D (2004) Service brands and communication effects. J Mark Commun 10(4):241–254
Oh H, Parks SC (1997) Customer satisfaction and service quality: a critical review of the literature and research implications for the hospitality industry. Hospitality Res J 20:35–64
Oke A (2007) Innovation types and innovation management practices in service companies. Int J Oper Prod Manag 27(6):564–587
Oke A, Idiagbon-Oke M (2010) Communication channels, innovation tasks and NPD project outcomes in innovation-driven horizontal networks. J Oper Manag 28:442–453
Orfila-Sintes F, Crespí-Cladera R, Martínez-Ros E (2005) Innovation activity in the hotel industry: evidence from Balearic Islands. Tour Manag 26(6):851–865
Ottenbacher M, Shaw V, Lockwood A (2006) An investigation of the factors affecting innovation performance in chain and independent hotels. J Q Assur Hospitality Tourism 6(3–4):113–128
Pappu R, Quester PG, Cooksey RW (2005) Consumer-based brand equity: improving the measurement—empirical evidence. J Prod Brand Manag 14(3):143–154
Pendergast D (2010) Getting to know the Y generation. Tourism Gener Y 1:1–15
Phillips-Wren G, Hoskisson A (2015) An analytical journey towards big data. J Decis Syst 24(1):87–102
Pikkemaat B, Peters M (2006) Towards the measurement of innovation—a pilot study in the small and medium sized hotel industry. J Q Assur Hospitality Tourism 6(3–4):89–112
Pine BJ, Peppers D, & Rogers M (2009) Do you want to keep your customers forever? Harvard Business Press, Boston, MA
Pivčević S, Garbin Praničević D (2012) Innovation activity in the hotel sector—the case of Croatia. Ekonomska istraživanja 1:337–363
Potts J, Mandeville T (2007) Toward an evolutionary theory of innovation and growth in the service economy. Prometheus 25(2):147–159
Prensky M (2001) Digital natives, digital immigrants part 1. On Horiz 9(5):1–6
Rababah K, Mohd H, Ibrahim H (2011) Customer relationship management (CRM) processes from theory to practice: the pre-implementation plan of CRM system. Int J e-Education e-Business e-Management e-Learning 1(1):22
Rahimi R (2017) Customer relationship management (people, process and technology) and organisational culture in hotels: which traits matter? Int J Contemp Hospitality Manag 29(5):1380–1402
Reid M (2002) Building strong brands through the management of integrated marketing communications. Int J Wine Mark 14(3):37–52
Reinartz W, Dellaert B, Krafft M, Kumar V, Varadarajan R (2011) Retailing innovations in a globalizing retail market environment. J Retail 87:S53–S66
Rezaei S (2015) Segmenting consumer decision-making styles (CDMS) toward marketing practice: a partial least squares (PLS) path modeling approach. J Retail Consum Serv 22:1–15
Ringle CM, Wende S, Becker J (2014) SmartPLS 3, SmartPLS, Hamburg. Available at https://www.smartpls.com/. Accessed 22 Jan 2018
Ritter T, Walter A (2012) More is not always better: the impact of relationship functions on customer-perceived relationship value. Ind Mark Manage 41(1):136–144
Ruiz-Molina ME, Gil-Saura I, Moliner-Velázquez B (2011) Does technology make a difference? Evidence from Spanish hotels. Serv Bus 5(1):1–12
Ruiz-Molina ME, Gil-Saura I, Servera-Francés D (2017) Innovation as a key to strengthen the effect of relationship benefits on loyalty in retailing. J Serv Mark 31(2):131–141
Ryssel R, Ritter T, Gemunden HG (2004) The impact of information technology deployment on trust, commitment and value creation in business relationships. J Bus Ind Mark 19(3):197–207
Salegna GJ, Goodwin SA (2005) Consumer loyalty to service providers: an integrated conceptual model. J Consum Satisfaction Dissatisfaction Complaining Behav 18:51–67
Sansone M, Colamatteo A (2017) Trends and dynamics in retail industry: focus on relational proximity. Int Bus Res 10(2):169

Santos-Vijande ML, López-Sánchez JA, Rudd J (2016) Frontline employees' collaboration in industrial service innovation: routes of co-creation's effects on new service performance. J Acad Mark Sci 44(3):350–375
Sasmita J, Mohd Suki N (2015) Young consumers' insights on brand equity: effects of brand association, brand loyalty, brand awareness, and brand image. Int J Retail Distrib Manag 43(3):276–292
Schultz DE (1999) Integrated marketing communications and how it relates to traditional media advertising. In: Jones JP (ed) The advertising business: operations, creativity, media planning, integrated communications. Sage, London, pp 325–338
Scupola A (2014) The relation between innovation sources and ICT roles in facility management organizations. J Facil Manag 12(4):368–381
Sigala M, Connolly D (2004) In search of the next big thing: IT issues and trends facing the hospitality industry. A review of the sixth annual pan-european hospitality technology exhibition and conference (EURHOTEC 2001); International Hotel and Restaurant Association, Paris, 19–21 February, Palais Des Congres. Tourism Management, vol 25, issue no 6, pp 807–809
Sigala M (2011) eCRM2.0 applications and trends: the use and perceptions of Greek tourism firms of social networks. Comput Hum Behav 27:655–661
Sigala M (2012) Web 2.0 and customer involvement in new service development: a framework, cases and implications in tourism. In: Sigala N, Christou E, Gretzel U (eds) Web 2.0 in travel, tourism and hospitality: theory, practice and cases. Ashgate Publishing Limited, London, pp 25–38
Sigala M (2015) The application and impact of gamification funware on trip planning and experiences. Electron Markets 25(3):189–209
Sigala M (2016) Social CRM capabilities and readiness: findings from Greek tourism firms. Information and communication technologies in Tourism 2016. Springer, Cham, pp 309–322
Sigala M, Marinidis D (2012) Web map services in tourism: a framework exploring the organisational transformations and implications on business operations and models. Int J Bus Inf Syst 9(4):415–434
Smerecnik KR, Andersen PA (2011) The diffusion of environmental sustainability innovations in North American hotels and ski resorts. J Sustain Tourism 19(2):171–196
So KKG, King C (2010) When experience matters: building and measuring hotel brand equity. The customers' perspective. Int J Contemp Hospitality Manag 22(5):589–608
Srivastava RK, Fahey L, Christensen HK (2001) The resource-based view and marketing: the role of market-based assets in gaining competitive advantage. J Manag 27(6):777–802
Srivastava RK, Shervani TA, Fahey L (1998) Market-based assets and shareholder value: a framework for analysis. J Mark 62(1):2–18
Strutton D, Taylor DG, Thompson K (2011) Investigating generational differences in e-WOM behaviours. Int J Advertising 30(4):559–586
Sundbo J, Orfila-Sintes F, Sørensen F (2007) The innovative behaviour of tourism firms: comparative studies of Denmark and Spain. Res Policy 36(1):88–106
Tantiseneepong N, Gorton M, White J (2012) Evaluating responses to celebrity endorsements using projective techniques. Q Market Res Int J 15(1):57–69
Tejada P, Moreno P (2013) Patterns of innovation in tourism 'small and medium-size enterprises'. Serv Ind J 33(7–8):749–758
Thwaites D, Lowe B, Monkhouse LL, Barnes BR (2012) The impact of negative publicity on celebrity ad endorsements. Psychol Mark 29(9):663–673
Tourespaña (2017) Accommodation in Spain. Available at http://www.spain.info/en/informacion-practica/alojamientos/. Accessed 22 Jan 2018
UNIDO (2002) Innovation and learning in global value chains. In Industrial development reports 2002/2003 (105–115). Available at http://www.unido.org/fileadmin/user_media/Publications/Research_and_statistics/Branch_publications/Research_and_Policy/Files/Reports/Flagship_Reports/IDR/Industrial%20Development%20Report%202002-2003.pdf
Vargo SL, Lusch RF (2004) Evolving to a new dominant logic for marketing. J Mark 68(1):1–17
Vargo SL, Lusch RF (2008) Service-dominant logic: continuing the evolution. J Acad Mark Sci 36(1):1–10

Vila M, Enz C, Costa G (2012) Innovative practices in the Spanish hotel industry. Cornell Hospitality Q 53(1):75–85

Vize R, Coughlan J, Kennedy A, Ellis-Chadwick F (2013) Technology readiness in a B2B Online retail context: an examination of antecedents and outcomes. Ind Mark Manage 42(6):909–918

Volo S (2006) A consumer-based measurement of tourism innovation. J Q Assur Hospitality Tourism 6(3–4):73–87

Wang CJ, Tsai CY (2014) Managing innovation and creativity in organizations: an empirical study of service industries in Taiwan. Serv Bus 8(2):313–335

Wang H, Wei Y, Yu C (2008) Global brand equity model: combining customer-based with product-market outcome approaches. J Product Brand Manag 17(5):305–316

Wang D, Xiang Z, Fesenmaier DR (2014) Adapting to the mobile world: a model of smartphone use. Ann Tourism Res 48:11–26

Wang D, Xiang Z, Fesenmaier DR (2016) Smartphone use in everyday life and travel. J Travel Res 55(1):52–63

Weerawardena J (2003) The role of marketing capability in innovation-based competitive strategy. J Strateg Mark 11(1):15–35

Weill P, Ross J (2009) IT savvy: what top executives must know to go from pain to gain. Harvard Business School Publishing, Boston

Williams KC, Page RA, Petrosky AR, Hernandez EH (2010) Multi-generational marketing: descriptions, characteristics, lifestyles, and attitudes. J Appl Bus Econ 11(2):21

Wong IA (2013) Exploring customer equity and the role of service experience in the casino service encounter. Int J Hospitality Manag 32:91–101

Wong MH (2015) Japan's robot hotel opens its (automatic) doors. CNN. Available at http://edition.cnn.com/2015/07/17/travel/japan-hotel-robot-opens/

Wu F, Yeniyurt S, Kim D, Cavusgil ST (2006) The impact of information technology on supply chain capabilities and firm performance: a resource-based view. Ind Mark Manage 35:493–504

Xiang Z, Schwartz Z, Gerdes JH, Uysal M (2015) What can big data and text analytics tell us about hotel guest experience and satisfaction? Int J Hospitality Manag 44:120–130

Ye J, Marinova D, Singh J (2012) Bottom-up learning in marketing frontlines: conceptualization, processes, and consequences. J Acad Mark Sci 40(6):821–844

Yoo B, Donthu N (2001) Developing and validating a multidimensional consumer-based brand equity scale. J Bus Res 52(1):1–14

Yoo B, Donthu N, Lee S (2000) An examination of selected marketing mix elements and brand equity. J Acad Mark Sci 28(2):195–211

Chapter 13
Big Data and Its Supporting Elements: Implications for Tourism and Hospitality Marketing

Mine Inanc–Demir and Metin Kozak

Abstract Big data plays a catalytic role in the determination of consumers' preferences while achieving meaningful results together by obtaining the right data. Artificial intelligence systems, particularly those powered by machine technology, can achieve significant results through the rapid elimination of large data sets. This leads to determining structural changes both in consumer behaviour models and marketing strategies. With preliminary information about consumers, intelligent system mechanisms, i.e. artificial intelligence and Internet of Things (IoT), have increased the speed of information processing and the analysis of larger volumes of information and have also targeted reaching the right consumer segments, affecting their decision-making preferences before the event. As a result, these types of automations may enable tourism and hospitality businesses to benefit from marketing activities with the help of different algorithmic solutions. Thus, this chapter aims to debate how big data, artificial intelligence and IoT are likely to reshape the traditional structure of tourism and hospitality marketing in the future and introduces new approaches as the key elements in maintaining competitiveness in a new era.

Keywords Big data · Artificial intelligence · Internet of things · IoT · Decision-making · Tourist behaviour · Tourism marketing

13.1 Introduction

Information and communication technologies (ICTs) have created an extensive market that means both suppliers and consumers benefit from their practical applications in our daily lives. For instance, as of January 2018, the number of Internet users has reached 4.0 billion, there are 5.1 billion mobile users and active media users 3.1, of whom 2.9 billion have access to social media through mobile devices (Global Digital Report 2018). Specifically, as of 2016, almost half of companies have invested in using big data (48%). ICT expenses will reach 2.7 trillion USD in 2020, includ-

M. Inanc–Demir · M. Kozak (✉)
Dokuz Eylul University, İzmir, Turkey
e-mail: M.Kozak@superonline.com

M. Sigala et al. (eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality*,
https://doi.org/10.1007/978-981-13-6339-9_13

ing telecommunications, services, cloud, mobility, smartphones, consumer IT and storage. Big data technology is expected to reach 58.9 billion USD by 2020 (International Data Corporation 2018). In addition, according to the study, called "Digital Universe", completed by the International Data Corporation (Turner and Gantz 2014), digital data will double every two years and by 2020 the amount of data will reach 44 zetabytes (44 trillion gigabytes). In addition, the number of "things" or devices connected to the Internet, at 13.4 billion now, will reach 38.5 billion in 2020 due to the influence of artificial intelligence and big data on people's behaviour in understanding the world.[1] These figures also show a dramatic increase in the rate of storage, circulation and usage of the amount in use.

With the development of ICT as a sign of the transition to the era of information and telecommunication in the new millennium, the result has been the restructuring of everyday life practices of human-beings and the uninterrupted flow of information. New tools have emerged to facilitate the typical daily lives of human beings worldwide by expanding the limits of data usage to a wider population and introducing the sky/space as the new forms of data storage. Among these are big data, machine learning, Internet of Things (IoT), and artificial intelligence (AI), as interrelated subjects (O'Leary 2013). As a direct consequence, there is the appearance of unlimited data that can be transacted not only within the world but also between the world and space. This also brings transformations in understanding the changing needs of consumers in the new era.

The production of large quantities of data in the digital era is as important in the field of communications/new media as in every single field. The analysis of big data and data mining in social media is one of the methods adopted when considering the speed and quantity of data flow. Furthermore, computer-based systems such as AI have great prospects in facilitating the collection and organization of data via digital media. The volume of data achieved by using AI constitutes an important position in analysing big data. Unlike any forms of other data, current computer systems allow data to be stored with a very large capacity. For example; TV channels, newspapers, chain businesses and transport companies collect and store data about millions of people, particularly regarding their demographic characteristics, experiences and expectations. This data can be collected through the websites of businesses or with the help of intermediary agencies. As is known, with the beginning of the post-modern era, people have begun to individualize and emphasize the formations on this side. While it was possible to reach people through mass communication channels (e.g. newspapers, TV, radio, etc.) during periods when AI had not yet been developed, it is now possible to access different communication channels with different contents for each individual by using more personalized marketing instruments such as social media and e-mails.

What is stated above is also expected to lead to determining structural changes in consumer behaviour models (Song and Liu 2017). Having the preliminary informa-

[1] 'Internet of Things' connected devices almost triple to over 38 billion units by 2020. https://www.juniperresearch.com/press/press-releases/iot-connected-devices-to-triple-to-38-bn-by-2020. Accessed on 3 March 2018.

tion about consumers, intelligent system mechanisms, i.e. AI and IoT, have processed the speed of information and the analysis of growing volume and have also aimed at reaching the right consumer segments affecting their decision-making preferences beforehand. As a result, these types of automations may enable the tourism and hospitality businesses to benefit from marketing activities with the help of different algorithmic solutions. Thus, this chapter aims at debating how big data, AI and IoT are likely to reshape the traditional structure of tourism and hospitality marketing in the future by giving examples from the practice and introducing the new approaches as the key elements in maintaining the competitiveness of new era.

13.2 Big Data and Its Supporting Elements

Today, developments such as iCloud, social media, big data and IoTs are shaped under the name of new ICTs, with great opportunities offered by technology, unlimited capacity and ability to emerge. Such new systems, with their interactive and dynamic structure, enable consumers to produce and share data on a variety of activities in a network environment, and to experiment with what they consume. Parahalad and Ramaswamy (2002) consider this transformation as a "co-creation of value". Tourists are able to transform their holiday consumption into experiences by using these intelligent system applications (Stienmetz 2018). As long as they feel pleased with their holiday experiences, they may like to share these moments on new media and any pleasure obtained through such posting such as positive comments by friends and gaining likes helps tourists maximize their holiday experiences.

By continuously analysing big data, the production/distribution processes are adjusted according to the situation. In order to increase productivity with the analysis obtained, big data brings new institutional practices to the tourism industry. As indicated by O'Leary (2013), these new institutional practices, as a part of big data, are important in a different dimension together with the emergence of the IoT. The IoT creates a global network where each object and each individual communicate with each other. This dispersed global network operates as a system that is open and accessible at all times. Such technologies, equipped with wireless sensor networks, provide incredible contributions to collecting large data sets. The feedback that will be obtained with big data, the automation systems and the skilled algorithms will provide a valuable contribution to the tourism industry. For instance, estimations for 2020 include 10% of the data produced on digital platforms being derived from machines and objects that can be connected to the Internet.

Due to the rapid change in data collection methods, AI has quickly gained momentum in favour of facilitating the collection and organization of data in digital media and has gained momentum in the last few decades. The volume of data and the aggregation speed achieved with AI has reached an important position in the analysis of big data. Just like big data, AI also leads to increase the volume, variety and speed of data. In the case of larger data volume, AI is useful in recognizing and learning other computer-based approaches that are difficult to understand. For example, various contents produced on social media platforms are detected by methods such as sentiment analysis through AI programs. In the tourism context (Stienmetz 2018; Zach et al. 2018), this helps to estimate tourists' destination preferences. Hence, AI provides new approaches that can create businesses' own marketing strategies.

Machine learning is a kind of AI that computers learn from the data. Machine learning is widely used in AI applications. Businesses like Facebook and Google take the advantage of using machine learning in estimation and data mining. All these applications are used to analyse data relating to the recognition of voice and face, and also the translation and classification of texts. Recently, there have been very rapid changes in the practice of machine learning. Developments in new learning algorithms have become a major issue in the application of AI, where the cost of machine learning becomes cheaper and ultimately reaches a large amount of data. Not only in science, technology and commercial areas but we can also see its applications in decision-making processes in the fields of health, education, finance, security and marketing.

13.3 Implications for Tourism and Hospitality Marketing

Generally speaking, the literature has accommodated quite a large number of studies on big data, AI and IoT and its unlimited implications for various forms of marketing products and services (e.g., Wang et al. 2016; Scharl et al. 2017; Zach et al. 2018; Xiang et al. 2015). Specifically speaking, on the other hand, the field of tourism and hospitality has been limited in the consideration of similar subjects (Fuchs et al. 2014; Xiang et al. 2015). Of these, Fuchs et al. (2014) have looked at the ways in which big data analytics plays an important role in knowledge generation for tourist destinations. Xiang et al. (2015) investigated the meaning of big data and text analytics in understanding the guest experience and satisfaction with hospitality businesses.

Furthermore, there has been the evidence of empirical studies carried out for marketing purposes in various sub-sectors of the tourism industry. Of these, Höpken et al. (2015) analysed large quantities of data on hotel reservations and consumer feedbacks to provide the authorities with support for decisions and maintain the optimization of tourism products and services. Menner et al. (2016), with the support of sentiment analysis, tried to explore the influence of consumers' opinions of tourism services and tourist destinations on the potential consumer decisions to book their reservations. In a similar study by Park et al. (2016), the objective included the influence of Twitter messages on the choices and intentions of cruise passengers. In a more destination-based vacation experience, Gong et al. (2016) aimed at exploring tourists' purpose of visit and their travel patterns while at a destination through the analysis of routes followed by taxi services.

However, all the studies carried out to date are no longer up to date as the process moves much faster and the meaning of big data has become the quantity of data counted with millions of users or words. From the business point of view, big data technology has already led organizations to evolve from decision-making methods to data-driven decisions (Lisi and Esposito 2015). Many organizations will be able to use this innovation to solve problems faster and easier and to better understand existing problems. Thus, machine learning and AI will play a key role in making organizations use the data to hand more intelligently and translate them into many non-structured interpretations. Computers and AI applications are the result of such services (Lisi and Esposito 2015).

Organizations with a lower potential for AI will be unaware of developments in marketing opportunities. Those organizations using AI are expected to attract more consumers and increase their daily marketing productivity. As a result of the big data-AI solidarity being established, the analysis of consumer needs will be made more effective and more comprehensive and the most appropriate product groups will be provided with the help of computers. Below is a list of implications of using big data and its derivations for maintaining effective up-to-date marketing programmes for tourism, travel and hospitality businesses in the context of four marketing mix elements (4Ps).

13.3.1 Product

The structure of tourism products is likely to become more intangible as the technology appears to play a greater role in the market. As such, the combination of big data and machine learning is expected to help creating new tourism products and services.[2] As a simple example, when the tourist or passenger is watching a film on a plane or in a hotel room, they can ask for more information about the location such as the list of major tourist attractions, weather forecasts, availability of transportation, and possibility of booking travel to the destination. The system automatically produces a single offer combining media, entertainment, travelling or taking vacations. Such a practice appears to have an indirect influence over promoting certain attractions or destinations and also helps to upgrade their values.

Big data is also useful for maintaining an effective tourist flow management (Ahas et al. 2008; Li and Yang 2017; Önder et al. 2016). Like maintaining the visitors' quality of stay at hotels, visitor experience can also be improved by controlling the human traffic in queues at airports and tourist attractions at the destination level. For instance, the city of Barcelona is able to measure how long visitors stayed in the area, whether they actually entered the *Sagrada Familia* cathedral, and the busiest times for visiting.[3] The local government was able to control the tourist flow and public services such as the deployment of security forces or arranging extra bus services considering the busiest times of visiting. In the future, the system may also allow the

[2]Defining the future of travel through intelligence. A discussion paper from Amadeus, 2016, Amadeus IT Group.

[3]Barcelona IoT, big data projects help manage tourists at popular attractions. https://internetofbusiness.com/barcelona-iot-big-data-tourist/. Accessed on 19 February 2018.

calculation of tourists' spending on a daily and/or attraction basis while on a holiday through the records of their self-reports or of credit cards using IoT.

In addition to wearable technologies such as smart watches, provision of data, especially through cameras or sensors, plays a primary role in interacting with IoT. Based on this evidence, it is now possible to find out how tourists travel, how they are connected in daily activities and how they contribute to their personal experience (Xiang et al. 2015). With positioning technologies, it is possible to determine tourist routes, the distances covered, and the profile of tourists who visit again from GPS location information. Thus, such data can help to determine tourist behaviours at different scales, to identify travel activities and to provide services for tourists' possible plans that can be regarded as the new forms of product development.

13.3.2 Promotion

One of the best examples of collaboration between big data and AI is the introduction of promotions in certain destinations, locations or facilities, particularly after the placements are made through social media. Which of our friends have had the personal experience, the contents of their reviews or the advertisements of various businesses may appear in our social media account. All these procedures are managed by AI created in a digital environment, not by human beings. In this way, it is possible to prepare specific information packages for each person. For example, when a person has finished searching for a destination, hotel, airplane or car rental by browsing different search or reservation engines, in a few seconds the social media account of this person starts promoting the same or different destination alternatives.

As for the individual perspective, as the individual's life continues and their preferences increase, the content of private personalized communication channels has also become richer (Song and Liu 2017). The AI platform accommodates news, products or information packs chosen via a list of specific key words. The keywords may include the person's political views, social status, age, gender, education level, hobbies, consumption patterns or experiences etc. Potential consumers can also benefit from these services, which are offered to them in alignment with their lifestyle and preferences, easily and cheaply through computers or smartphones. Thus, the marketing style has been achieved by meeting consumers' individual preferences, and even direct marketing techniques have been gaining importance, which defies the attractiveness of mass marketing techniques which aim to sell the uniform products and services but emphasizes the fact that each individual has different needs.

With the help of transaction-based and transactional data, it is possible to provide a snapshot by providing information about the contents of economic relations between service providers and consumers in the tourism industry (Scaglione et al. 2016). Approximately 80% of the data used in tourism constitute big data such as search records, shared social media contents, location information, visual materials (videos/photographs), sensor data, GPS signals and traffic movement on visits (Wang et al. 2016). For instance, airline and hotel businesses are also among the first to create and take advantage of such tools by improving the quality of their loyalty programmes. There are the practices of distributing personalized information to consumers based on their loyalty status and the footprints of their past search behaviours (Davenport 2013). If one's Facebook friends have recently been to a certain attractive destination, a flight or hotel business can offer a certain discount for the person to try it. Or if a consumer may have not been to the same restaurant for a long time, the business may offer them a free drink for their next visit in a week. In case a passenger is likely to miss the flight, the airline business may send a message to book the next flight with a small percentage of penalty.

13.3.3 Price

The terms of both price and value are interconnected. Any possible change in the former or later can in/directly influence the other from the perspective of consumers. In today's information era, developments in technologies add value to products and services. Being the first in the market may sometimes contribute to brand image that makes to strengthen its value. Moreover, the system can also automatically safeguard the value even when the price is not stable. As a result of the combination between big data and AI, airline and hotel businesses have successfully pioneered the use of price optimization analytics. According to the findings of an industry report by Amadeus (Davenport 2013), Air France and KLM, for instance, keep passengers' data for two years. The system automatically makes calculations and optimization of the revenue for origin and destination itineraries and identifies price levels based on passengers' profiles. The system further estimates the possible rate of cancellation and no-shows on flights that assists in estimating the rate of overbooking to be used.

Software programs such as Hadoop and MapReduce can also be used for the analysis of big data. Hadoop is an open source software that analyses everything from server logs to GPS signals, Twitter to email and sensor reading, storing all the large volumes of structured and unstructured data. This provides a data processing environment that can be scaled to complex and large data volumes (Davenport 2014). Therefore, through the analysis of big data, tourism and hospitality businesses will have a lot of effects that increase profitability and efficiency as a result of correct data processing and will reshape their marketing strategies. For instance, with more than 150,000 cases, Scaglione et al. (2018) have concluded that the determinants of last minute behaviour include the country of origin, season, length of the stay, composition of the party and destination. This may help the businesses reposition itself based on the revisited pricing strategies.

13.3.4 Place

The predictions indicate that the connected things will reach 38.5 billion by 2020.[4] Such an incredible progress is also likely to influence the tourism system in the near future. As the core element of this system, place has now been seemingly replaced by a digitalised platform that maintains all procedures in an online platform. The new place has become more transparent offering the potential consumers more opportunities to collect more information prior to their purchasing by comparing more alternatives in the choice set. Such a dynamic platform makes businesses to keep their eyes open for playing an active role in the game of competitiveness.

In the consumption stage, such changes will also allow businesses to create new techniques for their consumers to improve the quality of their service experience. Hotel guests do not need to wait at the front desk as they will be automatically informed once their rooms are ready to move in. Mobile keys will help them open the room. These keys will be equipped with the personal data of hotel guests. Using a digital identity, visitors can book their rooms directly using their mobile applications. The identity will collect visitors' preferences to offer them a more personalized experience for their next visit. When they go into the room, the TV screen will log into the e-mail account to show the list of new incoming messages. Visitors will also be helped to regulate their room temperature, see the list of TV channels based on the frequency of their previous likes, and decide the type of minibar items that have been most often consumed.[5]

In sum, from planning to experiencing vacations, data are being released by consumers in large quantities during various stages such as travel, entertainment, accommodation and restaurant services. These data reflect the people's real actions, not based on the consequences of survey data (Song and Liu 2017). Consumers themselves generate outward-facing information, leaving digital traces, especially with social-based services such as social media. All kinds of intelligent systems and social media provided by IoT make a big difference in terms of consumer marketing by being used in tourism and hospitality businesses. It brings not only the sense of marketing/consumers, but also the ability to solve problems and uncertainties in terms of the management of facilities.

[4] 'Internet of Things connected devices almost triple to over 38 billion units by 2020. https://www.juniperresearch.com/press/press-releases/iot-connected-devices-to-triple-to-38-bn-by-2020. Accessed on 3 March 2018.

[5] Tourism and The Internet of Things—IoT. https://medium.com/3baysover-tourism-networking/tourism-and-the-internet-of-things-iot-e41b125e7ddd. Accessed on 19 February 2018.

13.4 Conclusion

On the demand side, there has been a massive transformation in consumer behaviour. Consumers have become more experienced, independent and irrational due to changes in their values, lifestyles and demographic patterns. This has forced the supply side to shift from mass marketing to personalized marketing through the rules of market segmentation. The production process has also become more consumer-centric. Furthermore, as a sign of a new era in ICTs, concepts such as big data, IoT and AI have recently gained significant importance in many sectors because the developments in ICTs have accelerated worldwide. Tourism is one of the important fields that use these concepts and will also be influenced to a great extent.

Such changes will occur in the "P"s of tourism and hospitality marketing on the supply side and consumer behaviour on the demand side. As for the retransformation of tourism and hospitality marketing, new forms of "P"s can be explained as below: First, tourism products and services will be redesigned with the help of ICTs. Products and services will become more destination-oriented and smart destinations will be the core of tourism products and services. Second, the ability to use information technology and develop more technology-oriented products and services will be an indicator of pricing and value. Third, the place where all purchasing and transactions to be handled will become much more virtual. Finally, the promotion will also be more virtual-centric, where consumer decision-making can be influenced by the experience of other consumer peers and more personalized communication channels will be part of online or virtual marketing.

The next implications are more tourist behaviour-oriented. As repeatedly emphasized in this chapter, the combination of big data and other forms of technological advances will allow service providers in the tourism industry to create completely new products and services that will also make the visitors' life or vacation patterns easier. Big data plays a catalytic role in determining consumers' preferences while achieving meaningful results together with obtaining the right data. AI systems, particularly those powered by machine technology, can achieve significant results through the rapid elimination of large data sets. As in the case of "chicken and eggs", consumers will be automatically influenced by the data created by themselves that may force them to follow the crowd or trend.

As a result, it will be more likely to see consumers in tourism and hospitality become more flexible and irrational. This means that they may book their vacations and immediately be involved in the vacation experience with or without having the need to do so and being more keen on switching to other alternative products, services or destinations while still on a vacation. Moreover, the distance between the stages of consumer behaviour models appears to be very narrow or even almost zero, e.g. booking something even without any need being aroused, booking it as soon as the decision is made, starting consumption as soon as it is booked, completing the review about the holiday before it ends and so on. Updating tourism and hospitality marketing strategies will be an influential factor in such a transformation of consumer behaviour.

References

Ahas R, Aasa A, Roose A, Mark U, Silm S (2008) Evaluating passive mobile positioning data for tourism surveys: an Estonian case study. Tour Manag 29(3):469–486

Davenport TH (2013) At the Big Data crossroads: turning towards a smarter travel experience. Amadeus IT Group

Davenport TH (2014) Big Data @ Wor. Türk Hava Yolları Yayınları, Istanbul

Fuchs M, Höpken W, Lexhagen M (2014) Big data analytics for knowledge generation in tourism destinations: a case from Sweden. J Destination Mark Manag 3(4):198–209

Global Digital Report (2018) Digital in 2018: world's internet users pass the 4 billion mark. https://wearesocial.com/blog/2018/01/global-digital-report-2018. Accessed on 5 Mar 2018

Gong L, Liu X, Wu L, Liu Y (2016) Inferring trip purposes and uncovering travel patterns from taxi trajectory data. Cartography and Geogr Inf Sci 43(2):103–114

Höpken W, Fuchs M, Keil D, Lexhagen M (2015) Business intelligence for cross-process knowledge extraction at tourism destinations. Inf Technol Tourism 15(2):101–130

International Data Corporation (2018) Worldwide Big Data technology and services forecast 2016–2020. Accessed on 6 Feb 2018. https://www.idc.com/getdoc.jsp?containerId=US40803116

Li D, Yang Y (2017) GIS monitoring of traveler flows—based on Big Data. In: Xiang Z, Fesenmaier DR (eds) Analytics in smart tourism design concepts and methods. Springer, Cham, pp 111–126

Lisi FA, Esposito F (2015) An AI application to integrated tourism planning. In: Gavanelli M et al (eds) I*IA 2015, LNAI 9336. Springer, Cham, pp 246–259

Menner T, Höpken W, Fuchs M, Lexhagen M (2016) Topic detection: identifying relevant topics in tourism reviews, information and communication technologies in tourism 2016. In: Proceedings of the international conference, Bilbao, Spain, pp 411–423

O'Leary DE (2013) Big Data2, the 'Internet of Things', and the 'Internet of Signs'. Intell Sys Account, Finance and Manag 20:53–65

Önder I, Koerbitz W, Hubmann-Haidvogel A (2016) Tracing tourists by their digital footprints: the case of Austria. J Travel Res 55(5):566–573

Parahalad CK, Ramaswamy V (2002) The future of competition. Harvard Business School Press, Boston M.A

Park SB, Ok CM, Chae BK (2016) Using Twitter data for cruise tourism marketing and research. J Travel Tourism Mark 33(6):885–898

Scaglione M, Johnson C, Favre P (2018) When "last minute" really is "last minute". In: Stangl B, Pesonen J (eds) Information and communication technologies. Springer, Cham, pp 501–514

Scaglione M, Favre P, Trabichet J.-P (2016) Using mobile data and strategic tourism flows: pilot study MoniTour in Switzerland. In: Proceedings of the big data & business intelligence in the travel & tourism domain workshop, Östersund, Sweden, 11–12 Apr

Scharl A, Lalicic L, Onder I (2017) Tourism intelligence and visual media analytics for destination management organizations. In: Xiang Z and Fesenmaier DR (eds) Analytics in smart tourism design concepts and methods (pp 165–178). Springer, Cham

Song H, Liu H (2017) Predicting tourist demand using Big Data. In: Xiang Z, Fesenmaier DR (eds) Analytics in smart tourism design concepts and methods. Springer, Cham, pp 13–28

Stienmetz JL (2018) Reconstructing visitor experiences: structure and sentiment. In: Stangl B, Pesonen J (eds) Information and communication technologies. Springer, Cham, pp 489–500

Turner V, Gantz JF (2014) The digital universe of opportunities: rich data and the increasing value of the Internet of Things. http://www.emc.com/leadership/digital-universe/index.htm. Accessed on 8 Feb 2018

Wang D, Xiang Z, Fesenmaier DR (2016) Smartphone use in everyday life and travel. J Travel Res 55(1):52–63

Xiang Z, Schwartz Z, Gerdes J, Uysal M (2015) What can big data and text analytics tell us about hotel guest experience and satisfaction? Int J Hospitality Manag 44:120–130

Zach FJ, Wallace SA, Tussyadiah IP, Narayana SP (2018) Developing and testing a domain-specific lexical dictionary for travel talk on Twitter (#ttot). In: Stangl B, Pesonen J (eds) Information and communication technologies. Springer, Cham, pp 528–539