



Research paper

Tourism analytics with massive user-generated content: A case study of Barcelona

Estela Marine-Roig^{a,*}, Salvador Anton Clavé^b^a Department of Business Administration and Economic Management of Natural Resources (AEGERN), University of Lleida (UDL), C/ Jaume II, 73, Campus de Capponet, 25001 Lleida, Catalonia, Spain^b Rovira i Virgili University, Catalonia, Spain

ARTICLE INFO

Article history:

Received 6 April 2015

Received in revised form

17 June 2015

Accepted 18 June 2015

Available online 29 July 2015

Keywords:

User-generated content

Smart city

Smart tourism destination

Big data

Business intelligence

Catalonia

ABSTRACT

The aim of this paper is to highlight the usefulness of big data analytics to support smart destinations by studying the online image of Barcelona (a leading smart city and tourist destination) as transmitted via social media through the analysis of more than 100,000 relevant travel blogs and online travel reviews (OTRs) written in English by tourists who have visited the city in the last 10 years. The proposed methodology used in this paper facilitates the massive gathering, cleaning up, and analysis of tourism-related user-generated content (UGC) from the most suitable sources, and helps to define the transmitted image of the city through collecting and processing large volumes of digital data. It is also used to extract business intelligence (BI) from OTRs concerning visits to Barcelona's main landmark/attraction, La Sagrada Família. The findings of this massive content analysis of information from a trustworthy source, UGC data, is very useful in applying BI to destination management, both in order to develop and assess marketing strategies and to improve branding and positioning policies among tourism and marketing organizations. It reinforces the ability of cities such as Barcelona to develop a smart city and destination concept, as well as a strategy for themselves.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The concept of 'big data' refers to the massive accumulation of information and to the systems that manipulate these large datasets. Gandomi and Haider (2015) highlight the three Vs (volume, variety, and velocity) that characterize big data, and claim that traditional data-management systems are insufficient to manage it, giving rise to big data technologies capable of creating real-time intelligence from high volumes of various data. In this respect, Sanz (2013) asserts that cities with an appropriate operating system can store, analyse, and generate near-real-time business intelligence (BI) with big data collected from social media feeds, among other sources.

The spectacular growth of social media and user-generated content (UGC) on the Internet provides a huge quantity of information that allows for the firsthand ascertaining of the experiences, opinions, and feelings of tourism 'users' or customers (Marine-Roig & Anton Clavé, 2015; Xiang, Schwartz, & Uysal, 2015). The volume of data generated in social media has grown

from terabytes to petabytes (Gandomi & Haider, 2015; He & Chen, 2014), and data stored and analysed by big companies are set to move from the petabyte to exabyte magnitude soon (Hu, Wen, Chua, & Li, 2014). Social media can be classified into blogs, review sites, media sharing, question-and-answer sites, social bookmarking, social networking, social news, and wikis (Gandomi & Haider, 2015; Marine-Roig, 2014). Lu and Stepchenkova (2015) found that the main sources for studies on UGC data are, in order of frequency: consumer review websites, blogs, media-sharing websites, social networks, and virtual communities; the main topic areas being service quality, destination image and reputation, UGC as electronic word-of-mouth (eWOM), experiences and behaviour, and mobility patterns. In recent research, Koltringer and Dickinger (2015) have found that UGC is the richest and most diverse source of online information.

In the field of tourism, most authors agree on the importance of UGC (Koltringer & Dickinger, 2015; Lu & Stepchenkova, 2015; Marine-Roig, 2015) in the construction of destination image through the eWOM effect (Hidalgo, Sicilia, & Ruiz, 2014; Jalilvand, Samiei, Dini, & Manzari, 2012), and consider travel blogs, online travel reviews (OTRs), or online consumer reviews as rich sources of UGC data (Marine-Roig, 2014; Xiang et al., 2015). In the field of tourism and hospitality, a relative decrease in travel blogs can be observed, along side a tremendous growth in OTRs, especially in

* Corresponding author.

E-mail addresses: estela.marine@aegern.udl.cat (E. Marine-Roig), salvador.anton@urv.cat (S. Anton Clavé).

the hospitality sector (Marine-Roig & Anton Clavé, 2015). For instance, in January 2015 TripAdvisor asserted that it had reached more than 200 million reviews and opinions, Trivago had reached 140 million integrated user hotel reviews, Booking had collected 43 million verified reviews, and Expedia had gathered 11 million customer reviews. Given such figures, UGC should be identified as a valuable source of big data that is useful for the management of smart cities and smart tourism destinations.

This paper aims to highlight the usefulness of big data analytics to support smart tourism destinations by studying the online social media-transmitted image of Barcelona (a leading smart city and tourist destination) through the analysis of more than 100,000 relevant travel blogs and OTRs written in English by tourists who have visited the city in the last 10 years. To do so, it proposes a method of gathering and analysing big UGC data composed of five stages: destination choice, Web hosting selection, data collection, pre-processing, and content analysis. More specifically, a quantitative content analysis was conducted of 117,487 travel blogs and OTRs. This method is also used to extract BI from the 7481 OTRs on visits in 2014 to Barcelona's main landmark, La Sagrada Familia. The findings of this massive content analysis of information from a trustworthy source, UGC data, is of paramount usefulness in terms of applying BI to destination management, not only to develop marketing strategies but also improve branding and positioning policies among tourism and marketing organizations. It reinforces the ability of cities such as Barcelona to develop a smart city and destination concept, as well as a strategy for themselves.

2. State of the art

According to Del Chiappa and Baggio (2015), the concept of a smart tourism destination is still emerging and is arising from that of the smart city. A smart city is a city that performs in a forward-looking way in regard to six characteristics (economy, people, governance, mobility, environment, and living), and is built on the smart combination of endowments and activities of self-decisive, independent, and aware citizens (Giffinger et al., 2007). Boes, Buhalis, and Inversini (2015) indicate that a smart city focuses on its citizens, while a smart destination intends to improve tourist experiences through information and communication technologies (ICTs). They build a framework for the dimensions of the smart tourism destination that requires fundamental constructs (leadership, human capital, entrepreneurs, innovation, and social capital) supported by technology applications and a strong ICT infrastructure. This in turn, provides the basis to support the components of tourism (tourism experience, tourism competitiveness, and the six As of tourism: attractions, accessibility, amenities, available packages, activities, and ancillary services) and smart cities.

More specifically, Del Chiappa and Baggio (2015) define a smart tourism destination as a networked system of stakeholders delivering services to tourists, complemented by a technological infrastructure aimed at creating a digital environment that supports cooperation, knowledge sharing, and open innovation. In this vein, Buhalis and Amaranggana (2014) consider that 'smartness', when referring to a tourism destination, requires the dynamic interconnection of stakeholders through a platform capable of exchanging real-time information related to tourism activities, with the objective of maximizing user or customer satisfaction and resource management efficiency. These activities produce a large multidimensional set of digital information, which is understood within the concept of big data, and allows national tourism organizations (NTOs) and destination marketing organizations (DMOs) to extract valuable insights.

According to Wang, Li, and Li (2013), the use of big data by

smart tourism destinations can support business decision-making and optimal resource allocation, and can assist in the discovery of new insights in ways that affect markets and organizations. In their chapter on strengths for the tourism industry of using big data, Oliver et al. (2014) describe a set of advantages over traditional methodologies offered by analysing large amounts of data reliability, representativeness, information detail and segmentation capacity, the ability to 'hybridize' data with other current or future sources, new information flows, and the possibility of new business opportunities.

Hu et al. (2014) propose a definition for big data analytics based on the software, hardware, and aim of analysis: 'Big data analytics is the process of using analysis algorithms running on powerful supporting platforms to uncover potentials concealed in big data, such as hidden patterns or unknown correlations' (p. 656). They categorize the analysis process into two alternative paradigms: streaming and batch. Streaming processing is characterized by data being analysed as it arrives because near-real-time results are needed, such as in the case of online applications, and only a small part of the stream is temporarily stored in the memory. Conversely, batch processing is characterized by data first being stored, and then divided into chunks that are processed in parallel in a distributed system; finally, the intermediate results are aggregated.

Gandomi and Haider (2015) focus on big data analytical processes in two phases: data management (acquisition and recording; extraction, cleaning, and annotation; and integration, aggregation, and representation) and analytics (modelling and analysis, and interpretation). In the field of tourism destinations, Fuchs, Hopken, and Lexhagen (2014) propose a knowledge destination framework architecture that distinguishes between a knowledge creation layer (data sources, data extraction, data warehousing, and knowledge generation through data mining) and a knowledge application layer, where there is a destination management information system that grants stakeholders instant access to BI-based analysis results.

Lu and Stepchenkova (2015) observe that a growing number of UGC analytical works use specialized analytical and computational procedures to manage big data. However, in most studies, UGC data are manually collected; the manual handling of data is time-consuming, limits sample size, and facilitates researcher bias. These authors also found that, in general, methods to gather information are unclear, and that the technical details of data analyses are often incomplete. As an example of manually handling UGC data and limited sample size, in addition to the 122 cases gathered by Lu and Stepchenkova (2015) – He and Chen (2014), located 333 blog posts with Google Blog Search, filtered them manually, which left them with 317 relevant blog posts, using the so-called Blog Mining method, they then merge the posts in text files. Analyse these posts using the *Leximancer* programme allowed them to identify seven relevant themes.

Koltringer and Dickinger (2015) extracted destination brand identity and image from 5719 online documents through Web content mining and natural language processing. Their method was divided into the following stages: data gathering, keyword analysis, sentiment detection, category building, and correspondence analysis. As an example of the large-scale analysis of consumer-generated content, Xiang et al. (2015), analysed 60,648 online customer reviews from Expedia, corresponding to 10,537 hotels from the 100 largest cities in the United States. Using a web crawler, they gathered all available textual content for each city and each hotel, and then created a relational database with unique identifiers assigned to each hotel property, review, and unique word. Then, after a pre-processing phase, they analysed guest experiences.

In this respect, the massive analysis of UGC data is of great

Download English Version:

<https://daneshyari.com/en/article/1011040>

Download Persian Version:

<https://daneshyari.com/article/1011040>

[Daneshyari.com](https://daneshyari.com)