'You will like it!' using open data to predict tourists' response to a tourist attraction

Eleonora Pantano [a,1], Constantinos-Vasilios Priporas [a,*], Nikolaos Stylos [b]

[a] Department of Marketing, Branding & Tourism, Middlesex University Business School, London, United Kingdom
[b] Department for Marketing, Innovation, Leisure and Enterprise, University of Wolverhampton Business School, Arthur Storer Building, Nursery Street, Wolverhampton, City Campus Molineux (North), WV11AD, England, United Kingdom

HIGHLIGHTS

- Exploitation of open data to predict tourist's future preferences based on profile characteristics.
- The Random Forest method is employed to, first, train the system and, second, provide opt predictions and propositions.
- Better reach the target tourist markets, thus increasing the effectiveness of related marketing strategies.
- Applicable procedure for attractions and tourism destinations around the world.

ARTICLE INFO

ABSTRACT

The increasing amount of user-generated content spread via social networking services such as reviews, comments, and past experiences, has made a great deal of information available. Tourists can access this information to support their decision making process. This information is freely accessible online and generates so-called "open data". While many studies have investigated the effect of online reviews on tourists' decisions, none have directly investigated the extent to which open data analyses might predict tourists' response to a certain destination. To this end, our study contributes to the process of predicting tourists' future preferences via Mathematica^{TM}, software that analyzes a large set of the open data (i.e. tourists' reviews) that is freely available on tripadvisor. This is devised by generating the classification function and the best model for predicting the destination tourists would potentially select. The implications for the tourist industry are discussed in terms of research and practice.

1. Introduction

The recent advances in digital media technologies and environments, as well as the introduction and acceptance of sophisticated interactive software applications, have driven the digital evolution of marketing in the information society epoch (Garrigos-Simon, Lapiedra Alcamí & Barberá Ribera, 2012; Mekonnen, 2016). Digital social media has played a key role in this recently established sub-field of marketing and its rapid spread has transformed how information is accessed and shared (Di Noia, Mirizzi, Ostuni, Romito, & Zanker, 2012; Pantano, 2014). In particular, the impact of social networking sites (SNSs) on word-of mouth communications and decision making processes has been well reported (Chu & Kim, 2011; Fotiadis & Stylos, 2016; See-To & Ho, 2014). Digital marketers realize that to successfully attract and leverage the interest of SNS users, they need to increase the utility of social networks by offering value added services (Diffley, Kearns, Bennett, & Kawalek, 2011). Thus, SNSs are now expanding their capabilities by offering a diverse portfolio of build-in applications (apps) to meet social media users' needs for novel experiences (Tung, Jai, & Davis Burns, 2014); namely, customized topic-specific virtual spaces to better support user-generated content (UGC) (e.g. Facebook apps, YouTube), including reviews, comments on past experiences and recommendations for future purchases (Turban, King, Lee, Liang, & Turban, 2015). As researchers note, online reviews based on SNS users' profiles and established preferences are integral to formulating future preferences and affecting consumer purchases (Baka,

* Corresponding author. W107 Williams Building, Hendon Campus, The Burroughs, London, NW4 4BT, England, United Kingdom.
E-mail addresses: E.Pantano@mdx.ac.uk (E. Pantano), C.Priporas@mdx.ac.uk (C.-V. Priporas), n.stylos@wlv.ac.uk (N. Stylos).
[1] W109 Williams Building, Hendon Campus, The Burroughs, London NW4 4BT, England, UK.

2016; Chevalier & Mayzlin, 2006). The premise is that behavior is influenced not only by individuals' beliefs, feelings, impressions, and behavioral norms, but also by recommendations and prior experiences stemming from the social environment, which in turn produce attitudes and intentions (Cheng & Huang, 2013; Tsai & Bagozzi, 2014; White, 2005). In fact, the more the product online review features available to consumers, the higher the likelihood for sales of related items within the product category (Chevalier & Mayzlin, 2006).

Similarly, in a travel and tourism context, tourists' recommendations via tripadvisor, Yelp etc. influence other travelers' decisions about many different aspects of their trips, e.g. selection of a tourist destination, accommodation and attractions to visit (Filieri, Alguezaui, & McLeay, 2015; Hudson, 2014; Pantano & Di Pietro, 2013; Xiang, Magnini, & Fesenmaier, 2015). Notwithstanding the fact that some researchers have indicated that many reviews are fake, or overly positive or negative, consumers perceive online reviews as more trustworthy than content provided by official destination websites (Fotis, Buhalis, & Rossides, 2012). Drawing on a huge amount of UGC, marketers make systematic efforts to exploit as much open data as possible to support digital marketing effectiveness. These efforts could potentially improve online sales and the profitability of e-travel services (e.g., accommodation, transportation, restaurants, entertainment, sightseeing and tourism destination information) (Korfiatis, García-Bariocanal, & Sánchez-Alonso, 2012; Nguyen & Cao, 2015).

Nevertheless, up until now, most research in UGC and most online reviews have underlined the importance of analyzing ratings to increase the likelihood of travelers' having enjoyable trips (e.g. Fang, Ye, Kucukusta, & Law, 2016; Neidhardt, Seyfang, Schuster, & Werthner, 2015; Phillips, Zigan, Silva, & Schegg, 2015; Sotiriadis & van Zyl, 2013; Zhang, Zhang, & Yang, 2016), though only a few studies explicitly examine the impact of reviews on SNS users' future choices (i.e. Ayeh, Au, & Law, 2013; Jalilvand, Samiei, Dini, & Manzari, 2012; Pantano & Di Pietro, 2013; Sparks, Perkins, & Buckley, 2013). These studies focus on the readability, credibility and helpfulness of online reviews, however they do not explore the extent to which recommendations maybe perceived as useful to other travelers willing to travel to the same destinations (Schuckert, Liu & Law, 2015). Moreover, they do not investigate ways of using this data to improve traveller review sites' consultation capabilities to the benefit of individuals, hospitality businesses and tourist destinations at large.

Taken together, this study seeks to examine the extent to which open data analysis may apply to the tourists' process of selecting tourist destinations and/or services. In particular, we attempt to predict travelers' attitudes toward a tourist attraction by transforming large amounts of open data into value propositions. In doing so, we implement the random decision forest algorithm approach (Coussement & De Bock, 2013; Xie, Ngai & Ying, 2009) drawing on data available on a popular travelers' review site.

Given that few studies have explored the potential of open data to serve as means of providing opt vacations-related automated database-driven recommendations (Buhalis & Law, 2008; Gretzel, Sigala, Xiang, & Koo, 2015; Jannach, Zanker, & Fuchs, 2014), the objectives of the study are twofold. First, it aims to investigate the potential benefits of using open data sources to form appropriate future travel propositions, thus moving one step forward from the standard method of investigating the influence of perceived value as well as the reliability of online reviews on formulating intentions (Fang et al., 2016; Korfiatis et al., 2012; Lee, Law, & Murphy, 2011; Liu & Park, 2015; Sparks et al., 2013). Secondly, it seeks to highlight the effectiveness of leveraging a limited bulk of open data, as an alternative to big data sets, in terms of providing useful outputs.

From a theoretical point of view, this study draws attention to the huge potential of using online open data sources to influence tourists' attitudes and behaviors. Practically, we propose a computational tool that can greatly contribute to the effective positioning of hospitality organizations and tourist destinations.

2. Theoretical background

2.1. Open data

Open data has been defined by the Open Knowledge Foundation (http://okfn.org/) in 2005 as "data that can be freely used, shared and built on by anyone, anywhere, for any purpose". Maccani, Donnellan, and Helfert (2015) point out that there are 3 principles behind this definition: (1) availability and access (people can get the data); (2) re-use and redistribution (people can re-use and share the data); (3) universal participation (anyone can use the data). Furthermore, the volume of the information released through open data platforms is huge (Ojha, Jovanovic, & Giunchiglia, 2015; Wu, Liu, Chu, Chu, & Yu, 2014), it is based on a wealth of information and enables enhanced knowledge creation (Theocharis & Tsihritzis, 2013).

Kitchin (2014) asserts that the focus of open data could be any type of socio-economic or business phenomena but that in general, the emphasis to date has been on opening up data that has a high policy and commercial re-use value, such as economic, transport and spatial data. Today, open data are mostly provided by public and services providers (organizations, institutions, and enterprises) while the potential of open data for business development is still mostly unexplored (Pesonen & Lampi, n.d) For example, governments are trying to exploit open data to support the development of better services for citizens (Chan, 2013; Hielkama & Hongisto, 2013). Processing open data is recognized as a potentially powerful alternative to analyzing data collected via surveys (Gurstein, 2011). In specific, the use of open data is being increasingly acknowledged as a means of supporting knowledge management in various contemporary business and technological applications such as, smart cities (Inayatullah, 2011; Ojo, Curry, & Zeleti, 2015). Cities have been the first to be involved in processing open data in various applications (Longhi, Titz, & Viallis, 2014), such as the management of their tourist destination products (Buhalis & Amaranggana, 2013; Mariani, Buhalis, Longhi, & Vitouladiti, 2014), recognizing them as a key component of their smart city strategy (Marine-Roig & Clavé, 2015).

2.2. Open data in tourism

Tourism is by nature an industry in which marketing communications strongly depend on data exchange (Mack, Blose, & Pan, 2008). In today's rapidly changing world, various forms of data related to tourism activities and services are produced and utilized across a range of online applications (Buhalis & Law, 2008). This is primarily the outcome of the increasing ability to digitize growing volumes of data, and the development of open-sources and open data policies (Sabou, Braşoveanu, & Önder, 2015; Soualah-Alila, Coustaty, Rempulski, & Doucet, 2016). For tourist destinations there are significant opportunities to use open data to develop cultural sights, transportation, marketing and the environment (Wiggins & Crowston, 2011). As people have increasingly focused on the quality of the tourist experience, the demand for open data in tourism and hospitality research has becomes intense (Wu et al., 2014). A growing amount of tourism-related open data is now available on the platform in XML, CSV, or JSON format (Wu et al., 2014). According to Longhi et al. (2014) tourism is the first industry to be concerned with open data. Open data can facilitate local authorities in their planning processes (e.g., advertising) and in