What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management

Kun Kim [a], Oun-joung Park [b], Seunghyun Yun [a], Haejung Yun [c],*

[a] Yonsei University, Seoul, Republic of Korea
[b] Department of Tourism Management, Jeju National University, Jeju-si, Republic of Korea
[c] Department of International Office Administration, College of Science & Industry Convergence, Ewha Womans University, Seoul, Republic of Korea

A B S T R A C T

Recently, the Internet has brought a big change in tourists' behavior patterns. Travelers not only reserve hotels and airline tickets online, but also exchange travel information and descriptions of pleasant or unpleasant travel experiences through online review sites and personal travel blogs. In spite of the increasing use of online channels, application of online text data has been limited since the volume of the data set is too large to analyze manually and comprehensively. With recent technological advances in processing big data online, consumer-generated information can be automatically analyzed by artificial intelligence.

As an aspect of smart tourism, this study applied the sentiment analysis method to analyze travelers' online reviews of Paris. A total of 19,835 pieces of review data collected from a traveler review site (www.virtualtourist.com) were processed. All reviews were grouped into 14 categories as follows: overview, restaurants, sightseeing, hotels, things to do, night life, transportation, shopping, sporting & outdoors, favorites, off the beaten path, what to pack, tourist traps, warnings and danger, and local customs. Tourists' perception about the service in each category was successfully measured, and as an illustration, we chose "transportation" category that reported relatively low level of service quality for post-hoc analysis to reveal why tourists feel negatively about the transportation service.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

"The purpose of visiting and using websites has changed from read-only to read-write." (Cambria et al., 2013). This evolution has created enthusiastic users interacting with others and sharing information through social networks, online communities, blogs, wikis, and other collaborative media. Indeed, the Web has become a major communication channel. The large amount of information contained in microblogging web-sites makes them an attractive source of data for opinion mining and sentiment analysis (Cambria et al., 2013; Pak and Paroubek, 2010). Hence, many studies on mining online review data in the marketing and information business sector are in progress. Also the Web has encouraged a big change in tourists' behavior patterns. Travelers not only reserve hotels and airline tickets online, but also exchange travel information and descriptions of their pleasant or unpleasant travel experiences through online review sites and personal travel blogs as well.

In spite of the increasing use of online channels and contents, application of online text data in the context of destination hospitality services has been very limited because the volume of the data set is too large to analyze manually and comprehensively. However, with recent technological advances in processing online big data, consumer-generated online data can be automatically analyzed by artificial intelligence. In previous studies on tourism and hospitality services, most studies employed simple methods and targeted small quantity of reviews to understand customer's perceptions (Lee et al., 2011; Pan et al., 2007; Svetlana Stepchenkov, 2006; Tang et al., 2011). Currently, several studies used advanced text mining technique and analyzing large data set to get insights from online review data sets (e.g. Yee Liau and Pei Tan (2014) employed sentiment analysis to measure customer's sentiment level on airline service; Also, Mankad et al. (2016) in the hotel business contents, adopted similar tactics to get insights from user-generated content (UGC). Despite these efforts to apply the text mining approach, several limitations have been noted in existing tourism and hospitality studies:

First, the methods such as word counting, networks analysis can be useful to extract some important keywords; however, it is difficult to show positive or negative mood of the reviews. Similarly, the sentiment analysis can represent the degree of positivity or negativity of the data,

* Corresponding author.
  E-mail addresses: kapple0914@gmail.com (K. Kim), ounj.park@gmail.com (O. Park),
rogve22@gmail.com (S. Yun), yunhj@ewha.ac.kr (H. Yun).

but has little prescriptive and practical implications; in other words, using sentiment analysis, we can find out tourists perceive how negatively about certain destination services, but cannot discover why they feel like that. Second, in spite of the existing large volume of available data on travel review sites, the average number of words collected and analyzed in the previous studies was very small (around 100– 300 words), which was not able to generate sounding results.

In order to fill this current knowledge gap, the purpose of this research is to investigate visitors' perceptions of destination services through the hybrid analysis of travellers' online review data by employing both sentiment analysis methods for diagnosis and co-occurrence analysis for prescription. That is, our research question is that "What is the current sentiment level of tourists shown in their reviews on various designation services (e.g. hotels, restaurants, shopping, etc.) and why do they feel negatively about the certain services?" With the research findings, we suggest various action plans for improving the service performance of destination services. Furthermore, to overcome the methodological limitations of prior studies, we collected a vast amount of data (19,835 reviews) and employed hybrid analysis methods to derive more insight from the text data.

The results of this study provided valuable insights for hospitality and destination services marketer to understand perceptions and opinions of travellers visiting various destinations. Moreover, analysing customer-generated reviews is more economic and less time-consuming than a field survey and allows researchers to immediately and periodically estimate consumers' perceptual evaluation of service performance.

Next, we review the concept of big data, smart destination management, and related concepts, along with previous text mining research on tourism and hospitality. Then, we present research methodology, including research design and data collection. Following this, sentiment analysis and co-occurrence analysis, and their results are described. Finally, we discuss theoretical and practical implications based on research findings.

2. Literature reviews

2.1. Big data, user-generate-content (UGC) & smart destination management

"Data lies at the core of all smart tourism activities." (Gretzel et al., 2015a, 2015b) Recently, analysing user-generated-content (UGC) is getting much attention in tourism & hospitality industry. According to this trend, hospitality& tourism studies have been focusing on conceptualizing smart tourism and its ecosystem (Gretzel et al., 2015a, 2015b; Koo et al., 2013). Also, the researches on destination marketing research emphasize the importance of analysing UGC as an aspect of smart tourism (Buhalis and Amaranggana, 2015; Huang et al., 2012; Kim et al., 2014; Miguéns et al., 2008).

Buhalis and Amaranggana (2015) argued that "Smart Tourism Destinations should make an optimal use of big data by offering right services that suit users' preference at the right time". Kim et al. (2014) noted that

"Web 2.0 not only engages consumers but also increases the opportunities for businesses to interact with consumers." They proposed that DMO (Destination Marketing Organization) officer need to put more effort into analysing consumers' information processing on SNSs. Huang et al. (2012) also argued that to improve tourism service quality and promote tourism management, it is a necessary to combine the ICT with destination marketing.

Among big data which can provide valuable insights about the existing and potential customers, UGC especially has high value on tourism & hospitality marketing since user-generated data is more credible for potential customer's decision making than the host-generated data since tourism activities are the representative experience goods/services. (Akehurst, 2009; Cox et al., 2009; Miguéns et al., 2008; Ye et al., 2011) In this regard, previous research has proved that impact of UGC on tourist behavior & destination image. (Buhalis, 1998; Miguéns et al., 2008; Poon, 1993; Sheldon, 1997; Wang et al., 2002). As noted by Cox et al. (2009) a 10% increase in the ratings of user reviews can boost index of online hotel bookings, by more than 5%.

Although these attempts have highlighted the importance of analysis user-generated data, these are still in the preliminary and descriptive stage. Hence, the in-depth analysis of UGC, such as traveller's reviews might be one of the important key factors for successful destination marketing management.

2.2. Text mining on tourism & hospitality research

"Text mining, also known as text data mining or knowledge discovery from textual databases, refers generally to the process of extracting interesting and non-trivial patterns or knowledge from unstructured text documents" (Tan, 1999). Text mining not only for market surveys but also for various studies in diverse contexts (e.g. customer relationship management, biomedicine, tourism etc.) has been applied to previous research. Furthermore, several studies have employed text mining on various subjects in tourism & hospitality research.

As an aspect of restaurant management, research by He et al. (2013) applied text mining to analyse content on social media sites (e.g. Facebook and Twitter) of the three largest pizza chains: Pizza Hut, Domino's Pizza and Papa John's Pizza. The results revealed the value of social media competitive analysis and text mining as an effective technique to extract business value from the vast amount of available social media data. In the context of hotels, Bjorkelund et al. (2012) analysed textual reviews and visualized data on Google Maps, providing avenues for users to easily detect quality hotels. Lau et al. (2005) used text mining technology to analyze web documents of the Hong Kong Hotels Association and classified hotel types by counting words of web documents. Svetlana Stepchenkov (2006) conducted their research using text mining to analyze the destination image of Russia by counting words in reviews. Also, a study by Choi et al. (2007) applied a similar technique to the aforementioned study to analyze the destination image of travelers to Macau. (Tang et al., 2011) also applied a similar process and technique to conduct research in destination images.

Table 1
Text mining research on tourism & hospitality research.

Source	Data	Context	Finding	Methodology
Lau et al. (2005)	Web-documents	Hotel	Categorizing the types of hotels	Cluster
Svetlana Stepchenkov (2006)	Web-documents	Tourism (destination)	Destination image	Frequency analysis
Pan et al. (2007)	UGC (BLOG)	Tourism (destination)	Understanding of the destination image	Network analysis and content analysis method
Bjorkelund et al. (2012)	UGC (traveller's reviews)	Hotel	Reporting each hotel's level of sentiment, and marking it on the map	Sentiment analysis
He et al. (2013)	UGC (SNS)	Restaurant (pizza)	Competitive analysis and strategy	Content analysis
Duan et al. (2013)	UGC (traveller's reviews)	Hotel	Revisiting SERVQAUL by text mining algorithms	Text classification and sentiment analysis
Yee Liau and Pei Tan (2014)	UGC (traveller's reviews)	Airline	Service evaluation	Clustering and sentiment analysis