Technological Forecasting & Social Change 123 (2017) 362–369

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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 R T I C L E I N F O

Article history:
Received 5 April 2016
Received in revised form 6 December 2016
Accepted 2 January 2017
Available online 13 January 2017

Keywords:
Smart tourism
Smart destination management
Sentiment analysis
Text mining
User-generated content (UGC)

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.

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).

http://dx.doi.org/10.1016/j.techfore.2017.01.001

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

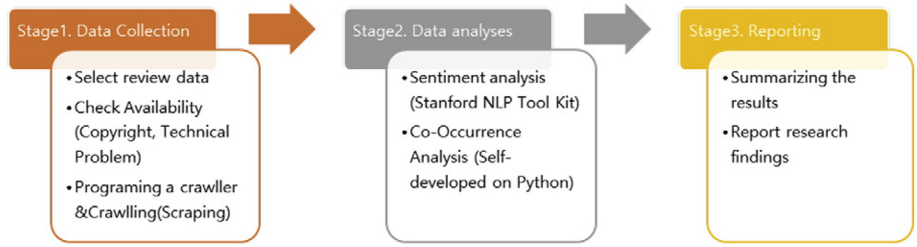

Fig. 1. Research process of UGC analysis on destination hospitality services.

Recently, using advanced technology, (Pan et al., 2007) used text mining to count the ratings of each positive and negative words on reviews to draw conclusions about service levels. In hospitality & tourism research, several researchers have employed text mining methods to analyse user-generated-reviews. For example, (Duan et al., 2013) used text mining to study service quality on hotels based on analysis of customers' online reviews; they identified factors that significantly impacted user service satisfaction. The representative text mining studies in the field of tourism & hospitality can be summarized as Table 1.

In order to extend these efforts, the big data analysis using more sophisticated and rigour techniques is required for smart destination management. For example, hybrid methods combining the existing techniques can generate more relevant implications.

3. Research methodology

3.1. Research design

The overall research process can be summarized as Fig. 1. In Stage 1, we select target data, collect, and process them for the text analysis. As a next step, in Stage 2, the data analyses which include sentiment analysis and post-hoc analysis (co-occurrence analysis) are done. In stage 3, we organize, summarize the results, and report research findings. The research process will be explained more in detail in the next section.

3.2. Data collection

Using the programming language Python3.3, this study collected anonymous travellers' online reviews of their experiences in Paris (Fig. 2) from the Virtual Tourist website (www.virtualtourist.com). We believe that Virtual Tourist is the most proper one for our research since it offers more various categories than other travel review sites; it has 14 categories (overview, restaurants, sightseeing, hotels, things to do, night life, transportation, shopping, sporting & outdoors, favourites, off the beaten path, what to pack, tourist traps, warnings and danger, and local customs), which could make DMO diagnose the current status of each destination services and prescribe detailed action plans.

With the web crawling program, reviews were extracted from the 14 default categories on the website (Table 2). As an early stage of big data analysis using text mining technique, we selected Paris as our research context since there are a large volume of cumulated customer reviews on Paris as one of the most representative tourism destinations.

4. Analysis and results

4.1. Sentiment analysis

Sentiment has been analysed as an important variable in marketing and is also considered a critical indicator of consumer behaviour. The positive influence of consumer sentiment on purchase intention has already been verified by existing research (Carroll et al., 1994; Gaski and Etzel, 1986; Mishkin et al., 1978; Throop, 1992).

In recent years, several computer science researches developing sentiment analysis algorithm are being conducted. There are several algorithms used in the sentiment analysis; in this study, we applied deep-learning[1] based on the Stanford sentiment analysis. The deep-learning-based sentiment analysis is latest and appropriate for analysis of syntactic structures and it is more accurate than others algorithms (Socher et al., 2013).

Table 3 shows the sample results of sentiment analysis. We have randomly extracted four samples on hotel categories, and two samples on transportation categories.

This study conducted sentiment analysis on reviews grouped into Virtual Tourist's 14 categories of travel experience in Paris. The 19,835 reviews were analysed by the Stanford sentiment analysis tool based on JAVA1.7.0_65. Scores were assigned from 4 to 0 (4: very positive, 3: positive, 2: neutral, 1: negative, 0: very negative.) The results are summarized in Table 4:

Based on the sentiment analysis, reviews on “night life”, “restaurants”, and “things to do” were found to be more positive than reviews in the other categories. This indicates that travellers seem to feel positive about their experiences associated with “restaurants”, “night life”, and “things to do in Paris”. The most negative reviews were categorized under “warning and danger” and “tourist traps”. These results indicate that travellers feel unpleasant about their experiences in those respective categories. As shown above, the sentiment analysis can act as a useful tool to find out the level of travellers' sentiments such as positive or negative mood. However, through sentiment analysis, we can only notice the level of customers' sentiment, and cannot find out why they feel positive or negative sentiment. Therefore, we attempted to perform co-occurrence analysis as a post-hoc analysis to identify causes of the problems identified.

4.2. Co-occurrence analysis

Previous hospitality studies adopting text mining approach have limitations in terms of their practical values. They only report sentiment scales of their data without explanations of what made customers feel negatively or positively. Considering that the ultimate goal of evaluating service quality is service performance improvement, identifying the causes of issues related to service is essential for sustainable business practices. As a solution, we employed co-occurrence analysis as a post-hoc analysis.

Co-occurrence analysis is a kind of content analysis technique that uses patterns of co-words of pairs of items (i.e., words or noun phrase) in a corpus of texts to identify relationships (He, 1999). Co-occurrence analysis is an important bibliometric way to map the relationship among concepts, ideas, and problems (Callon et al., 1983; Liu et al., 2011; Small, 1973). Co-occurrence analysis has mainly been used to

[1] Deep-learning, a remarkable technology by (Gartner, 2013) is a study of machine learning and is a means of managing human thoughts with computers. The deep-learning projected computer system functions similarly to a neural circuit in the brain and works as a human brain.

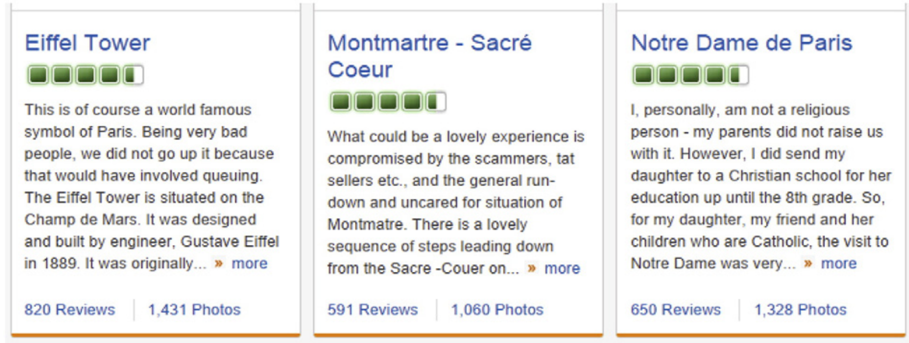

Fig. 2. Travel reviews (Virtual tourist).

analyse research trends or structure in information & library science studies. In various fields, researcher have adopted co-occurrence analysis to determine cognitive constructs of topics from text data. (Liu et al., 2011; Wang et al., 2015; Yan et al., 2015) As a result of the analysis, researchers can diagnose issues of their service.

Therefore, we performed co-occurrence analysis, as a post-hoc analysis, of "Transportation" category which showed a relatively low service performance in order to provide the insights about why customers feel negatively about the certain category.

Before conducting co-occurrence analysis, data pre-processing, such as stemming and lemmatization to make a common base form by reducing inflectional forms, was done by Stanford NLP (Manning et al., 2008). As a next step, co-occurrence analysis, using only the negative reviews of which sentiment scores are 0 and 1 points, was performed by Python3.3.

And the results are illustrated as Table 5. We reported the most frequently shown together four to five words with the three representative words (taxi, metro, and bus) in transportation category, after excluding some function words (e.g. "I", "we", and prepositions).

As a result of co-occurrence analysis, the term "expensive" appeared in the context of taxi and metro. The words, "dense" and "dirty", appeared in the association with "bus". Also in the metro review contexts, "luggage" appeared to be the high in the word frequency count. Because the French subway is very narrow in some sections, reviewers frequently complain about the insufficient space for baggage and the difficulties in carrying luggage when transferring to another subway line (e.g. "There are no lifts or escalators"). In the bus and taxi review contexts, drivers were frequently counted on the reviews. After manually checking the reviews containing the word "driver", we found out many cases of complaints related to the driver's demeanour and unfair billing practices (Refer to sample 5 and sample 6 of Table 3). Through this process, Paris's tourism board is able to not only recognize current problems but also evaluate the necessity of proposing solutions to solve these problems.

5. Conclusion and discussion

Although many industries pay great attention to analysing big data, studies of elements of the hospitality industry has been lagged.

Table 2
Review categories of Virtual tourist (Paris).

Category	# of reviews	Category	# of reviews
Overview	991	Local customs	677
Hotel	1174	Warnings and dangers	757
Restaurants	1558	Tourist traps	351
Things to Do	8809	What to pack	208
Night life	438	Off the beaten path	1174
Transportation	1448	Sporting & outdoors	78
Shopping	694	Favourites	1476

Recognizing the current knowledge gap, this study analysed reviews of travel experiences in Paris using sentiment analysis. The application of automatic sentiment analysis using artificial intelligence on destination tourism and hospitality services enables us to estimate tourist sentiments without undertaking economically challenging and time consuming research methods such as surveys, observation, and interviews in the field. Sentiment analysis is well suited to securing objectivity in consumer research, particularly in the analysis of the feelings and emotions of unsatisfied travellers, comparing with the manual content analysis done by human coders, since there are no inter-rater reliability problems. Also, it successfully reduces the cost and time necessary to collect and analyse data and to complete the analysis process.

We were able to measure the service performance in each category of travellers' experiences and compare the perception of service levels among 14 categories by analysing sentiment of customers on UGC. Among those categories, we chose "Transportation" category that reported relatively low level of service performance for post-hoc analysis to reveal why the reviewers feel negatively about the transportation service.

5.1. Theoretical contribution

This research has values on initial big data analysis research for smart tourism field. Considering that "Research in the area of smart tourism remains very limited and mostly provides case studies of existing initiatives"(Gretzel et al., 2015a, 2015b). In this regard, first of all, this study is significant in providing a step-by-step guideline for future research by demonstrating a hybrid text mining process. Also, this study could be applied into future smart tourism studies as a part of business intelligence, data analytics and other data-based research on UGC. Second, this study is the kind of interdisciplinary research, which represents currently popular research trend. Both hospitality & tourism management cases based on online reviews and computer science studies based on text analytic algorithms were successfully combined to derive more meaningful research outcomes.

5.2. Practical implications

Lately, significance and demand on big data analysis are increasing in the hospitality field. However, technical barriers and lack of knowledge on big data analysis prevent service firms from utilizing it. Whole process of this research could suggest guideline on the service and hospitality firms, which want to bring text analysis on their business.

First, to cope with the limitation of previous text mining studies we conducted hybrid text analysis on the user-generated-reviews. The results of this study not only reported numeric data (scales) of previous text mining research or Dashboard on BI Systems, also showed what traveller remarked on their service. According to this strength, marketer and service managers can get more insightful information on their service to boost service performance than before. For instance, this

Table 3
Samples of sentiment analysis (tourist's reviews on hotel & transportation in Paris).

Sample 1 – hotel (score = 3.5)	This was one of my favorite places to stay my entire tour around Europe. A lot of other reviews have this place marked as un-kept and un-clean, but I found it to have the best atmosphere of almost all the places I stayed, It has a bar on site and a courtyard right in the middle of it all, it's not a hostel where you sit in your room and read a book about Paris its somewhere you stay if you want to meet people and LIVE Paris.
Sample 2 – hotel (score = 2.25)	I stayed here with my friend several years ago and was repulsed. The place is dirty and full of American frat boys who are in Paris to get drunk and wake up the next day and do it all over again - in the courtyard, so every room can hear them all night. The rooms were dark and cold. The only thing good about the place is its location on lovely Commerce Street.
Sample 3 – hotel (score = 1.64)	This was a very quiet, relaxed hostel. Most rooms are single sex only with three bunk beds and shared bathrooms. There are a few rooms with six beds as well as a toilet and shower. They are the same price, approximately 21 euros per person. Unique Quality: The hostel is relatively near to the metro, and few changes are necessary to get to downtown. Rec room with internet. Very relaxed atmosphere. Maybe a little less "homey" than other hostels, but very clean. Unlike some other Parisian hostels, there are many rooms, so it is unlikely that it will be booked in advance, and they take reservations. Directions: From metro Porte de Bagnolet, several signs direct you to the hostel. It is a short 5 minute walk.
Sample 4 – hotel (score = 0.86)	A horrible experience – avoid at all costs!!! I arrived and was told that there were no rooms because they needed emergency repairs and that I'd be taken care of at another equivalent quality hotel. After schlepping on foot across town, I found they'd sent me to a fleabag hotel where my room was a closet next to the manager's desk on the ground floor. I hiked back to the hotel where magically a room was available but they wanted a credit card before they'd show me what it looked like. Check the reviews on tripadvisor.com – misuse of credit cards and this fraud with overselling rooms is apparently standard practice for them. While having it out with the clerk, another guest in the hotel came to complain that the hotel had taken and lost their passports – 'check back later, maybe we'll find them' was the response. Unique Quality: Rude staff who, to put it politely, stretch the truth
Sample 5 – transportation (score = 1.13)	My friend had arranged for a taxi to pick me up at his apartment to take me to where I had to catch the bus to CDG. I was down there about 10 min before the taxi was to arrive because I wanted to make sure I didn't miss it. Well a taxi pulled up about 15 min later and I assumed it was for me. The taxi driver said a name that was not mine and so I waited… He waited and his pickup did not show up. Then I asked him again who he was picking up, again, the name was not mine. I was getting worried and I had to get to the bus stop, so I decided to ask him if he could take me, he refused because I was not the one he came to pick up. Still his person did not show up. Worried I would miss my bus and miss my flight, I started walking towards the metro. Now this was about 4:30 in the morning, it was dark and I didn't want to go into the metro. As I walked away, the taxi driver spoke up and said he'd take me. I was so flustered and worried I'd miss the bus, I never noticed that he had the meter running all that time he was waiting. I noticed it about 1 min after we were driving. I'm pretty sure he was there to pick me up, but decided to take advantage of me and scam me. What an #$$! I was so pissed I was ready to just jump from the cab. I told him to stop because he had kept the meter running and I was not going to pay what was on the meter. I knew how much I should be paying and it was way, way past that. I argued with him and kept telling him to stop the cab, but he continued to drive. I just finally gave up because at that point we were almost there. He stopped, I got out and he wanted to be paid before he gave me my luggage, which was in the trunk. I paid him, he gave me my luggage and I told him he was a VERY BAD MAN and that what goes around comes around. I will NEVER take a taxi in Paris again. I'll walk or take the metro.
Sample 6 – transportation (score = 0.67)	Last Monday at the early hour of 5.30 am a taxi was booked to take me to Porte Maillot to catch the coach to Beauvais for a Ryanair flight (difficult enough at any time during the day…) Everything started to go terribly wrong when the booked taxi failed to show up. My pyjama clad boyfriend assisted me in trying to flag down an alternative taxi, after 10 min of trying a taxi stopped and, as I have now experienced an incalculable number of times, the driver could not be arsed taking me to my destination. The result: eventually there were simply no taxi about so 170 Euros had to be paid up for a new flight ticket (many thanks Juan), two meetings cancelled and a day at work missed. This seems to be the trend in Paris, for whatever their personal preference of passenger they have very negative/non progressive approach to customer service… I say bring in the European competition. I strongly advise anyone booking a taxi to catch a flight have secured provision, get a number and have a backup plan!! XXXX XXX TO FRENCH TAXI DRIVERS…

*Scale of score [4: very positive, 3: positive, 2: neutral, 1: negative, 0: very negative.]

research not only evaluated service performance of each service categories, also suggested detailed insight on transportation service by employing hybrid text mining approach. According to these results, destination marketing officer and government officer could understand issue on transportation service and need to improve it. And it could be distinguished competitiveness of Paris.

Second, considering that "Data lies at the core of all smart tourism activities." (Gretzel et al., 2015a, 2015b), it could expedite application to data analytics on smart tourism by guiding to business intelligence. For example, when destination marketing officer want to monitor their service performance in real time, they could apply customized analytics framework on their customer-relationship-management system.

Table 4
Results of sentiment analysis (Paris).

	Overview	Things to do	Favourites	Restaurants	Transportation	Hotel	Off the beaten path
Mean	1.8764	1.7602	1.6937	1.8296	1.5354	1.6851	1.6403
S.D.	0.6484	0.6031	0.5605	0.5459	0.5371	0.4770	0.5713
Unit	991	8810	1476	1558	1448	1174	1174
Rank	–	3	5	2	11	6	8

	Sporting & outdoors	Shopping	Local customs	Night life	Tourist traps	What to pack	Warning and danger
Mean	1.5809	1.7142	1.6430	1.9030	1.4911	1.5597	1.4061
S.D.	0.4513	0.5088	0.5313	0.5802	0.4420	0.4256	0.4009
Unit	78	694	677	438	352	208	757
Rank	9	4	7	1	12	10	13

# of reviews	Average	Standard deviation
19,835	1.6494	0.5104

(Note: S.D. means standard deviation.)

Table 5
Co-occurrence Analysis (Transportation).

	Words	Frequency		Words	Frequency
Taxi	Airport (+CDG[a])	54 (+28)	Metro	Airport (+CDG)	54 (+28)
	Driver	45		Not	24
	Hotel	34		Expensive	21
	Traffic	24		Luggage	18
	Expensive	21		Only	15
Bus	Airport	101	Bus	Dense	2
	Driver	31		Dirty	2

[a] CDG: IATA code of Paris main airport (Paris Charles de Gaulle Airport).

More practically, the several expected advantages of suggested research process on destination marketing can be summarized as follows:

Firstly, Applying sentiment analysis on destination marketing research could reduce research cost and time. Destination marketer could easily monitor on UGC, it makes the firms be always ready for the travellers and their issues on their service. Adopting census approaches in the field survey is expensive and time-consuming; however, all the customer-generated data can be relatively easily included in the analysis using our hybrid methodology; hence, sampling bias can be minimized. Once you develop a customized algorithms and set up automated procedures, it is relatively easy to monitor and analyse customer reviews almost at a real-time basis.

Secondly, as our research findings revealed, the hybrid-method could provide practical and detailed insights about reasons why travellers feel negatively about certain travel destinations. For example, transportation service providers could prioritize the improvement of service levels about the domains frequently mentioned in extremely negative reviews.

Last, but not least, the newly discovered knowledge and facts from analyzing vast amount of unstructured text such as UGC, could be shared and utilized by various stakeholders, such as hospitality service providers, policy makers, and educational institutes. Based on the findings, they have to develop employee training programs, tourism promotion policies, and new curriculums in order to make tourism and hospitality services and strategies really “smart”.

5.3. Limitations and concluding remarks

This research has the following limitations: First, Stanford Analysis Library Tool currently recognizes the English language only and this led to a problem where only English reviews were selected and processed using sentiment analysis. In future research, data analyses that support a diverse set of languages from different regions are necessary. Second, we investigated only one travel information website and only one destination, Paris, which is hard to be generalized. Expanding the amount of subject matter to encompass a broader set of data and also collecting data from other geographical regions is necessary. Third, this study adopts data-driven and exploratory manner by analysing large volume of tourists' review. Therefore, more theory-driven and confirmative approach and interpretation should be attempted in the near future.

Although several limitations exist as stated above, this research is significant as it is an initial study conducted using big data in the tourism field. More specifically, it is one of first studies to be conducted using a sentiment analysis based on artificial intelligence, as well as a post-hoc word co-occurrence analysis, that is applied to destination hospitality services. Also, this research deals with a larger set of data with practical analytical mechanisms compared to those of prior research studies. Thus, construction of know-how for future, large scale data-analysis is made available through this research. By improving the limitations of this research, this research is expected to give guideline for practical use and further study.

Appendix A

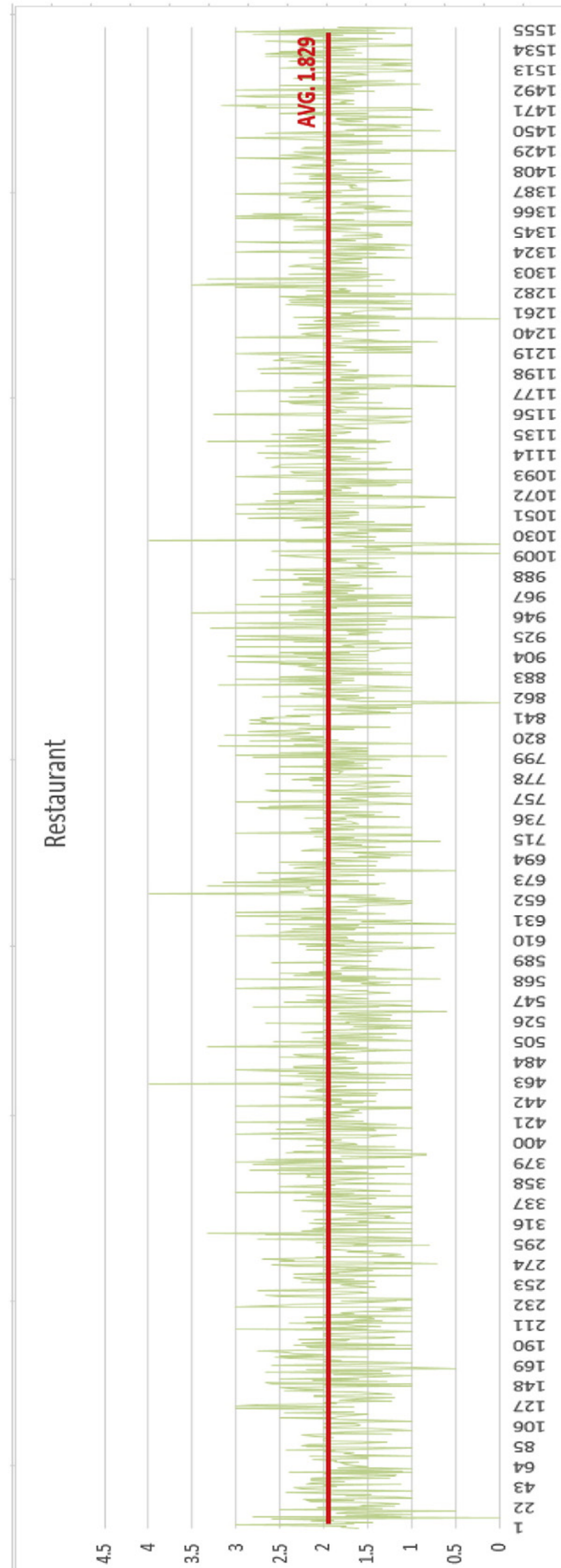

Fig. 3. Results of sentiment analysis (Restaurant).

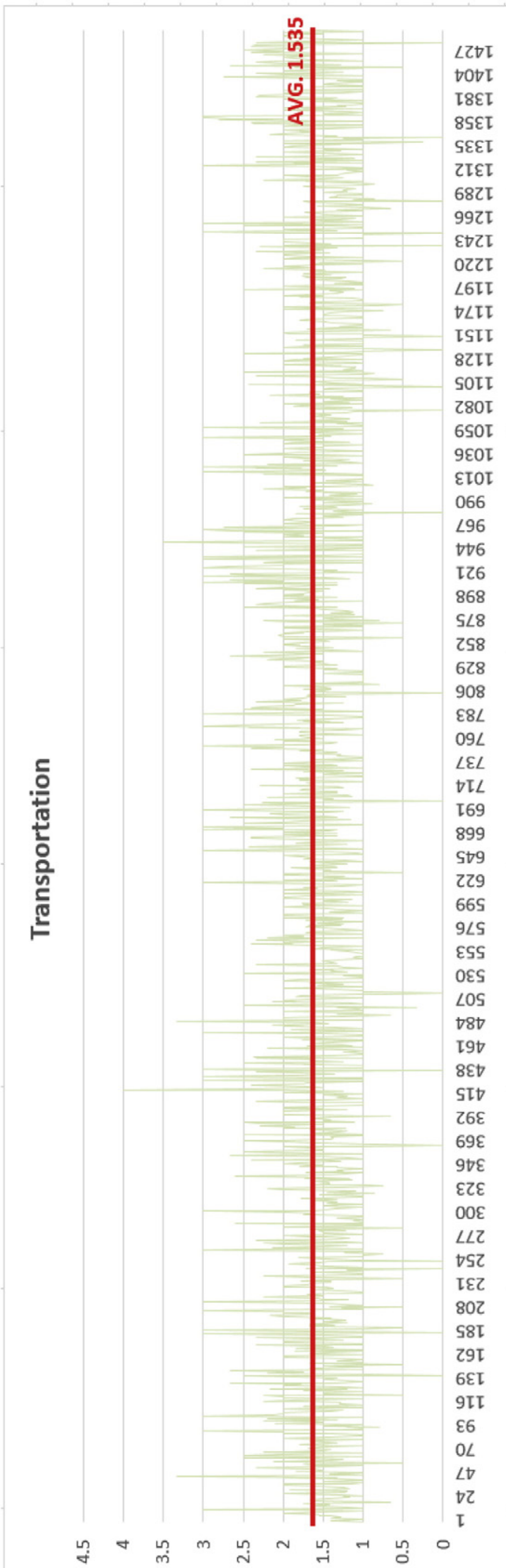

Fig. 4. Results of sentiment analysis (Transportation).

References

Akehurst, G., 2009. User generated content: the use of blogs for tourism organisations and tourism consumers. Serv. Bus. 3, 51–61.

Bjorkelund, E., Burnett, T.H., Norvag, K., 2012. A study of opinion mining and visualization of hotel reviews. Paper Presented at the Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services.

Buhalis, D., 1998. Strategic use of information technologies in the tourism industry. Tour. Manag. 19, 409–421.

Buhalis, D., Amaranggana, A., 2015. Smart tourism destinations enhancing tourism experience through personalisation of services. Information and Communication Technologies in Tourism 2015. Springer, pp. 377–389.

Callon, M., Courtial, J.-P., Turner, W.A., Bauin, S., 1983. From translations to problematic networks: an introduction to co-word analysis. Soc. Sci. Inf. 22, 191–235.

Cambria, E., Schuller, B., Xia, Y., Havasi, C., 2013. New avenues in opinion mining and sentiment analysis. IEEE Intell. Syst. 1.

Carroll, C.D., Fuhrer, J.C., Wilcox, D.W., 1994. Does consumer sentiment forecast household spending? If so, why? Am. Econ. Rev. 1397–1408.

Choi, S., Lehto, X.Y., Morrison, A.M., 2007. Destination image representation on the web: content analysis of Macau travel related websites. Tour. Manag. 28, 118–129.

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

Duan, W., Cao, Q., Yu, Y., Levy, S., 2013. Mining online user-generated content: using sentiment analysis technique to study hotel service quality. Paper Presented at the System Sciences (HICSS), 2013 46th Hawaii International Conference on.

Gartner, 2013. Gartner Reveals Top Predictions for IT Organizations and Users for 2014 and Beyond. Gartner Symposium.

Gaski, J.F., Etzel, M.J., 1986. The index of consumer sentiment toward marketing. J. Mark. 71–81.

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

Gretzel, U., Werthner, H., Koo, C., Lamsfus, C., 2015b. Conceptual foundations for understanding smart tourism ecosystems. Comput. Hum. Behav. 50, 558–563.

He, Q., 1999. Knowledge discovery through co-word analysis. Libr. Trends 48, 133–159.

He, W., Zha, S., Li, L., 2013. Social media competitive analysis and text mining: a case study in the pizza industry. Int. J. Inf. Manag. 33, 464–472.

Huang, X.K., Yuan, J.Z., Shi, M.Y., 2012. Condition and key issues analysis on the smarter tourism construction in China. Multimedia and Signal Processing. Springer, pp. 444–450.

Kim, S.-B., Kim, D.-Y., Wise, K., 2014. The effect of searching and surfing on recognition of destination images on Facebook pages. Comput. Hum. Behav. 30, 813–823.

Koo, C., Shin, S., Kim, K., Kim, C., Chung, N., 2013. Smart tourism of the Korea: a case study. Paper Presented at the PACIS.

Lau, K.-N., Lee, K.-H., Ho, Y., 2005. Text mining for the hotel industry. Cornell Hotel Restaur. Admin. Q. 46, 344–362.

Lee, M.J., Singh, N., Chan, E.S., 2011. Service failures and recovery actions in the hotel industry: a text-mining approach. J. Vacat. Mark. 17, 197–207.

Liu, G.-Y., Hu, J.-M., Wang, H.-L., 2011. A co-word analysis of digital library field in China. Scientometrics 91, 203–217.

Mankad, S., Han, H.S., Goh, J., Gavirneni, S., 2016. Understanding online hotel reviews through automated text analysis. Serv. Sci. 8, 124–138.

Manning, C.D., Raghavan, P., Schütze, H., 2008. Introduction to Information Retrieval. Cambridge University Press.

Miguéns, J., Baggio, R., Costa, C., 2008. Social media and tourism destinations: TripAdvisor case study. Adv. Tour. Res. 26, 1–6.

Mishkin, F.S., Hall, R., Shoven, J., Juster, T., Lovell, M., 1978. Consumer sentiment and spending on durable goods. Brookings Papers on Economic Activity, pp. 217–232.

Pak, A., Paroubek, P., 2010. Twitter as a corpus for sentiment analysis and opinion mining. Paper Presented at the LREC.

Pan, B., MacLaurin, T., Crotts, J.C., 2007. Travel blogs and the implications for destination marketing. J. Travel Res. 46, 35–45.

Poon, A., 1993. Tourism, Technology and Competitive Strategies. CAB International.

Sheldon, P.J., 1997. Tourism Information Technology. CAB International.

Small, H., 1973. Co-citation in the scientific literature: a new measure of the relationship between two documents. J. Am. Soc. Inf. Sci. 24, 265–269.

Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C., 2013. Recursive deep models for semantic compositionality over a sentiment treebank. Paper Presented at the Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).

Svetlana Stepchenkov, A.M.M., 2006. The destination image of Russia: from the online induced perspective. Tour. Manag. 27, 943–956.

Tan, A.H., 1999. Text mining: the state of the art and the challenges. Paper Presented at the Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery From Advanced Databases.

Tang, L., Tang, R., Scherer, A., 2011. Web site based destination images: a comparison of Macau and Hong Kong. J. China Tour. Res. 7, 2–19.

Throop, A.W., 1992. Consumer sentiment: its causes and effects. Fed. Reserv. Bank San Francisco Econ. Rev. 1, 35–59.

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

Wang, Z., Zhao, H., Wang, Y., 2015. Social networks in marketing research 2001–2014: a co-word analysis. Scientometrics 105, 65–82.

Yan, B.-N., Lee, T.-S., Lee, T.-P., 2015. Mapping the intellectual structure of the internet of things (IoT) field (2000–2014): a co-word analysis. Scientometrics 105, 1285–1300.

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, 634–639.

Yee Liau, B., Pei Tan, P., 2014. Gaining customer knowledge in low cost airlines through text mining. Ind. Manag. Data Syst. 114, 1344–1359.

Kun Kim received his Master degree (Business Analytics) from Yonsei University, Seoul. His research interests include Data analytics, Applied Statistics, Service Management, Text mining, and Social Media Strategy.

Oun-Joung Park is currently an Associate Professor of Tourism Management at Jeju National University in South Korea. Her recent research has focused on emotional aspects of tourist experience in hospitality and tourism industry.

Seunghyun Yun received Bachelor's degree (Electric & Electrical Engineering) from Yonsei University, Seoul. He is currently working as a Data Engineer at Yello Exchange, Korea. His research interests include Big Data Processing, AD-tech, Data analytics, Applied Statistics and Data & Text Mining. He has conducted projects on developing Cloud Systems for Big Data Processing & Analytics.

Haejung Yun is an Assistant Professor at Department of International Office Administration, College of Science & Industry Convergence, Ewha Womans University, Seoul, Korea. She received her Ph.D. degree in information systems from Graduate School of Information, Yonsei University. She previous worked for American University, Washington D.C. as a post-doctoral researcher. Her research interests include IT service management and information privacy. She has published papers in International Journal of Electronic Commerce, Journal of Electronic Commerce and Research, and other journals.