Contents lists available at ScienceDirect

Tourism Management

journal homepage: http://www.elsevier.com/locate/tourman

Automated topic modeling of tourist reviews: Does the Anna Karenina principle apply?

Andrei P. Kirilenko[a,*], Svetlana O. Stepchenkova[a], Xiangyi Dai[b]

[a] *Department of Tourism, Hospitality, and Event Management College of Health and Human Performance, University of Florida, P.O. Box 118208, Gainesville, FL, 32611-8208, USA*

[b] *College of Resource Environment and Tourism Capital Normal University, NO.105, North Road, The West 3rd Ring Road, Haidian District, Beijing City, 100048, PR China*

ARTICLE INFO

Keywords:
Traveler satisfaction
Latent Dirichlet allocation
Anna Karenina principle
TripAdvisor
Social media

ABSTRACT

Automated content analysis of online travel reviews allows identification of topics of travelers' satisfaction, yet its domain is not well researched. We suggest that the Anna Karenina principle positing a greater variability of the factors leading to business failure as opposed to those leading to success can be applied to the domain of visitors' reviews of historic and cultural attractions. The larger variability of issues in reviews of dissatisfied visitors is likely to result in limitations for automated topic modeling. We confirm our proposition using TripAdvisor reviews of the Terracotta Army museum in China, and validate the outcome with two additional sites. The study strongly suggests that application of unsupervised topic mining algorithms to negative reviews may be problematic and the results should be treated with caution. The main themes of dissatisfaction of visitors to all three sites are reported and practical implications for management of the attractions are discussed.

> "All happy families resemble one another; each unhappy family is unhappy in its own way."
>
> Leo Tolstoy, *Anna Karenina*

1. Introduction

Unlike durable or fast moving goods, tourism products and services cannot be physically displayed or inspected at the point of sale (Cho, 1998), thus tourists always try to find reliable and accurate information to reduce risks of buying (Jacobsen & Munar, 2012; Locander & Hermann, 1979; Maser & Weiermair, 1998). In the past, tourists often used recommendations of relatives and friends, while in the modern era they can resort to reviews and ratings from third party internet platforms such as TripAdvisor, Yelp, and others. Customer reviews are generally perceived as equally or more trustworthy than information found on official websites (Nielsen Global Online Consumer Survey, 2009). Moreover, online travel reviews often provide added value to tourists in the form of photos, videos, stories and location maps, which makes the information search engaging and convenient. It is widely accepted that online travel reviews play a significant role in the decision-making process of potential tourists (Mauri & Minazzi, 2013; Pan, MacLaurin, & Crotts, 2007). According to TripAdvisor, tourists have produced more than 830 million reviews and opinions in this platform (Tripadvisor, 2020), and 96% of their users, when booking a hotel, think it is important to read reviews, and 83% reference reviews for their decision (Feinstein, 2018).

The importance of online travel reviews is increasingly recognized by national and regional destination marketing organizations (DMOs) (Fuchs, Höpken, & Lexhagen, 2014; Stankov, Lazic, & Dragicevic, 2010; Stepchenkova & Zhan, 2013). Tourism reviews, however, are a type of big data and are intrinsically characterized by such features as volume, variability, velocity, and veracity. These characteristics refer to the scale of data, its various formats, the speed of data generation and processing, and "noisiness" or uncertainty. All these aspects of online reviews contribute to difficulty for tourism managers, marketers, and policy makers to find relevant content and analyze it in a timely and efficient manner (X. Li, Pan, Zhang, & Smith, 2009; Nonnecke, Andrews, & Preece, 2006; Schmunk, Höpken, Fuchs, & Lexhagen, 2013, pp. 253–265; Stepchenkova, Mills, & Jiang, 2007).

[*] Corresponding author. Department of Tourism, Hospitality, and Event Management College of Health and Human Performance, P.O. Box 118208, Gainesville, FL, 32611-8208, USA.

E-mail addresses: andrei.kirilenko@ufl.edu (A.P. Kirilenko), svetlana.step@ufl.edu (S.O. Stepchenkova), realnae@126.com (X. Dai).

https://doi.org/10.1016/j.tourman.2020.104241
Received 4 June 2020; Received in revised form 18 September 2020; Accepted 1 October 2020
Available online 13 October 2020

Since the overwhelming amount of online reviews are in an unstructured (plain text) or semi-structured (text and star-ratings) format, it is not feasible to manually extract and analyze them on a large scale. Meanwhile, with some programming skills, researchers can search for and download relevant information from review platforms, and data-mining algorithms are able to explore data structures in order to extract meaningful information with little assistance from the researchers (Raschka, 2015). The literature has shown great interest in applying data mining techniques to online travel reviews to study main topics in travelers' hotel evaluations, travel motivation, sentiments toward an attraction, and similar destination management problems (Y. Li, Ye, Zhang, & Wang, 2011; Rossetti, Stella, & Zanker, 2016; Xiang, Du, Ma, & Fan, 2017).

The methodological questions of automated topical analysis, however, have not been researched well (Newman, Lau, Grieser, & Baldwin, 2010). The primary obstacle for acceptance of topical analysis outside of the machine learning community involves the issues of interpretability of topics obtained with computerized algorithms (Mimno, Wallach, Talley, Leenders, & McCallum, 2011). Topical solutions that "mix unrelated or loosely related concepts substantially reduce users' confidence in the utility of such automated systems" (Mimno et al., 2011, p. 262). It has been estimated that "as many as 10% of topics may be so bad that they cannot be shown without reducing users' confidence" (ibid.) Given the recognition and managerial importance of online reviews, we examined the limitations of computerized text mining methods applied to online reviews of tourism and hospitality services, focusing specifically on the differences between positive and negative reviews as suggested by the Anna Karenina principle (TAK).

The overarching TAK principle states that while "no feature guarantees success, many guarantee failure" (Shugan, 2007). This was famously popularized by Jared Diamond (1999) in his bestselling book *Guns, Germs, and Steel: The Fates of Human societies*. Shugan (2007) states that one important outcome of the TAK is that "the most revealing variables might exhibit negligible variation among survivors because survivors are necessarily alike. Perhaps variability is inversely related to the variable's importance for survival" (p. 145). The TAK principle has been applied to many areas of science as diverse as statistics, geotectonic, pest control, and genetics. We found only one study in tourism that uses TAK: Tasci, Croes, and Villanueva (2014) cite the principle in connection to the rise and fall of community-based tourism. In relation to customer reviews, the difference in the nature of positive and negative reviews can be summarized as follows: The reviews of satisfied visitors are more similar than the reviews of dissatisfied visitors. In other words, variability of issues that lead to dissatisfaction is generally greater than those that are mentioned by satisfied reviews.

As applied to topic modeling, the variability of issues in negative reviews, together with the generally small number of negative reviews (Kladou & Mavragani, 2015; Lu & Stepchenkova, 2012) may affect topic extraction. Therefore, the main purpose of this study is to investigate whether automated methods of data mining are equally suitable for positive and negative online reviews to uncover topics of visitors' experience and, consequently, satisfaction and dissatisfaction with this experience. More specifically, we are set to demonstrate that standard, unsupervised data mining approaches may not fit well for negative reviews due to the nature of the data, i.e., the topical diversity of negative reviews combined with a smaller number of negative reviews as compared to positive reviews. In practice, this should result in less separability of dissatisfaction topics in the negative reviews. We also demonstrate that this phenomenon is stable across attractions in culturally different countries as well as languages of the visitors. In summary, this research investigates the following proposition:

Reviews of the customers with low satisfaction of attractions have low topic separability as compared to reviews left by satisfied customers, which results in poorly interpretable topics. This feature of negative reviews is invariant for culturally different attractions and visitors from different countries.

2. Literature review

2.1. Significance of negative reviews for tourism practice

The inseparability of production and consumption in tourism, high involvement and sophistication of modern long-haul leisure travelers, and the cost of travel relative to other living expenses elevate the potential for dissatisfaction. Relating these negative feelings on online review sites like TripAdvisor (Yoo & Gretzel, 2008) serves as a negative word-of-mouth (e-WOM) that affects destination image, desirability of attractions, and willingness to visit the attraction. Just as Amazon CEO Jeff Bezos once stated, "If you make customers unhappy in the physical world, they might each tell six friends. If you make customers unhappy on the Internet, they can each tell 6,000 friends" (Newman, 2015). Although positive reviews from tourists outnumber negative reviews (Kladou & Mavragani, 2015; Lu & Stepchenkova, 2012), negative information plays a more important role than positive information in consumers' buying decisions (Huang, Basu, & Hsu, 2010), because the former is generally considered more reliable (Ba & Pavlou, 2002) and useful to assess the quality of products (Casaló, Flavián, Guinalíu, & Ekinci, 2015; Herr, Kardes, & Kim, 1991). Combined with the high level of trust exhibited by customers toward online personal recommendations (65% of US adults who regularly checking online reviews believe them to be generally accurate (Smith & Anderson, 2016)), these negative reviews exert greater harm to a firm's reputation and financial performance. This is true for the hospitality and tourism domain as well (Sparks & Browning, 2011).

The effect of negative reviews is amplified by the star rating system on third party reviews platforms such as TripAdvisor, increasing their visibility for potential tourists. Thus, consumers' negative reviews have gained growing attention from both tourism businesses and academia (Yuksel, Kilinc, & Yuksel, 2006). From an attraction management viewpoint, monitoring online reviews and providing proper responses to dissatisfied customers is crucial for managers of destinations, tourism attractions (Olga & Raj, 2013), hotels (Dinçer & Alrawadieh, 2017; Zheng, Youn, & Kincaid, 2009), and restaurants (Pantelidis, 2010), being a sort of a managerial "virtual" service recovery instrument. Hospitality practitioners have also recognized the value of negative reviews in providing information on issues that need immediate managerial action with respect to product, service quality, and enhancing customer relations (Leung, Law, Van Hoof, & Buhalis, 2013).

2.2. Analysis of online reviews

The earliest published research in tourism and hospitality that used online reviews was a study by Harrison-Walker (2001) on online complaints against an airline. At that time, the analyses were largely conducted manually, and sample sizes typically ranged between 200 and 500 reviews (Au, Buhalis, & Law, 2014; Dinçer & Alrawadieh, 2017; Sparks & Browning, 2010; Zheng et al., 2009). Gradually sample sizes increased, driven by automation of data collection and wider awareness of computer-assisted text analysis methods between tourism academics and practitioners. A systematic literature review by W. Lu and Stepchenkova (2015, p. 141) reports that out of 122 papers that drew on user-generated content for tourism and hospitality applications, "24, 38, and 18 studies employed specialized software for data collection, data analysis, or both, respectively."

Nevertheless, the analysis of main topics of online reviews, such as the issues of satisfaction and dissatisfaction with the quality of hospitality product and tourism experience, remained primarily manual until recently. For example, Zheng et al. (2009) manually classified customer complaints about hotels into five categories, namely, rooms, service (with eight sub-categories), value, cleanliness, and dining. Au, Buhalis, and Law (2009) manually classified complaint reviews into nine categories: service, space, cleanliness, utilities/amenities, bedding, price, provision of amenities, decor, and miscellaneous. More recently, Dinçer and Alrawadieh (2017) used manual classification of 1- and 2-star reviews and reported that the most frequent e-complaints were about service quality, the efficiency of hotel facilities, and cleanliness and hygiene. Notably, these studies operated with datasets of fewer than 1000 reviews which may be considered a reasonable upper limit for the majority of academic studies employing manual content analysis. While limited data sets and interpretive qualitative methodologies can provide "fine granularity" and extensive details of analysis, the findings of studies that heavily rely on researchers' expertise in classification and interpretation are difficult to replicate (Hu, Zhang, Gao, & Bose, 2019).

The use of text mining algorithms in tourism and hospitality research has gradually increased, employing large data sets and more advanced computerized tools (Kirilenko, Stepchenkova, Kim, & Li, 2018; Liu, Teichert, Rossi, Li, & Hu, 2017; Marine-Roig & Clavé, 2015; (Xiang, Schwartz, Gerdes, & Uysal, 2015)). In the current literature, two main directions of large corpora of textual data have been clearly distinguished: (1) sentiment analysis (or polarity analysis) that focuses on extraction of positive, negative, or neutral sentiment expressed in a review for a particular attraction, hotel, or destination and (2) topical analysis (or topic modeling), which aims to extract the review's meaning. For sentiment analysis of textual data, a number of supervised and unsupervised analytical algorithms exist and have been applied to user-generated content; applicability of some of these tools has been evaluated for different types of tourism data (Kirilenko et al., 2018). In this paper, we focus on the topical analysis of online reviews.

2.3. *Extracting topics of online reviews with Latent Dirichlet Allocation*

The goal of topical analysis is to assist in understanding, classifying, and generalizing the meaning of a collection of documents (a corpus). While the first developments of automated topical analysis can be traced back to text indexing research in the early 1960s (Borko & Bernick, 1963), very few tourism papers used topic modeling prior to 2016 and all of them were published in non-tourism journals. It is only as recent as 2019 since the leading tourism journals have published more than a handful of papers that used topic modeling. Apparently, this development is owed to a recent crop of user-friendly software based on the Latent Dirichlet Allocation (LDA) method (Blei, Ng, & Jordan, 2003).

Recently, LDA applications in tourism have gained popularity: e.g., two major tourism journals, *Tourism Management* and *Journal of Travel Research* published two LDA-based papers in 2017 (Guo, Barnes, & Jia, 2017; Xiang, Du, Ma, & Fan, 2017), none in 2018, and 12 in 2019 (the third major journal, *Annals of Tourism Research*, published none). In comparison, another popular in-text mining topic analysis method based on the singular vector decomposition (SVD) was used only in three articles published in these tourism journals. Meanwhile, the applicability and limitations of the LDA method are largely unknown (Tang, Meng, Nguyen, Mei, & Zhang, 2014).

The experimental research of LDA-based topic modeling of multiple datasets (Tang et al., 2014) suggests the following informal guidelines: (1) the number of documents should be sufficiently large; (2) the length of documents should be large enough; (3) when either the number of documents or their length are above a certain threshold, topical analysis obtained from a sample are similar to the ones obtained from the entire corpus; (4) extracting overly large number of topics should be avoided; and (5) for LDA success, the topics should be well-separated, that is, the topics should be concentrated in a small number of words. It is not entirely clear if the practice of LDA analysis applied to negative reviews is following those guidelines and hence capable of returning coherent topics.

3. Method

To address the study's purpose, detailed analysis was performed on data collected from TripAdvisor reviews of the Museum of Qin Terracotta Warriors and Horses popular known as the Terracotta Army in Xi'an, China. The museum exhibits approximately 8000 life-size figurines of the imperial guards of Qin Shi Huang, the first emperor of the unified China. The excavated tomb is believed to be the "eighth wonder of the world" and one of the most significant archeological findings of the 20th century (Gülker, Hinsch, & Kraft, 2001). The Mausoleum of the First Qin Emperor is a China's State Priority Protected Site, listed in 1987 within the first group of world heritage sites (UNESCO, 2020). The historically significant site has attracted over 100 million visitors since its opening 40 years ago, and it has been repeatedly placed in the top 25 of TripAdvisor's "Traveler's Choice" list of the world's best museums, ranking first in China and Asia (Lu, 2019). However, the site's popularity also brings with it numerous problems related to transportation, crowdedness, logistics, food, and crime.

To validate the universality of the study's proposition, the same analytical approach was applied to the reviews of Red Square in Moscow, Russia, and Chichen Itza, Mexico (see Section 4.3). The additional sites were selected using the following criteria: (1) located at a destination geographically and culturally different from the main study area; (2) designated as a UNESCO World Heritage site; and (3) included in the TripAdvisor's Top Attractions list. Russia is the ninth most visited country in Europe with 24.6 million annual visitors (UNWTO, 2019), and Mexico is the most popular country destination in Latin America (and the second in North America) with 41.5 million annual visitors (UNWTO, 2019). Red Square and Chichen Itza are on the TripAdvisor Top Attractions list in their respective countries and both are included into the UNESCO World Heritage site list (https://whc.unesco.org). Thus, the criteria ensured that both attractions were of the same caliber as the Terracotta Army in their historic and cultural importance and were suitable for validation.

To answer the question of whether the positive and negative reviews are different in nature and, as such, lend themselves differently to probability-based statistical analysis methods like LDA, the following research steps were formulated in the study.

1. To test the invariance of the TAK principle, we first divided the reviews into two "cultural" categories: an English corpus and a Chinese corpus. Next, reviews were assigned to positive and negative categories based on their star rating. The subsets of reviews were compared with respect to their volume, star-rating distributions, and review length. The number of reviews and their distribution is of concern for the statistical solution, while the length of reviews is important from both the statistical and interpretational aspects.
2. The optimum number of topics was determined using both statistical indicators (perplexity and coherence) in combination with the "elbow criterion," as well as general interpretability and topic separation (Tang et al., 2014). In addition, previous research was consulted regarding the "optimum" number; it seems to range from 9 to 15 (Newman, Noh, Talley, Karimi, & Baldwin, 2010).
3. Topic interpretability is often decided "holistically," that is, by reviewing the generated list of top words defining a latent topic for the purpose of finding the best name describing the content of the topic, possibly aided by reading through the documents with the highest loading on the topic (examples can be found at Maier et al., 2018). To show that the difference in interpretability between positive and negative reviews is due to the nature of reviews and not the small sample size of the negative reviews, an additional analysis of a random subsample of positive reviews equal in size to the negative sample of reviews was also conducted.
4. The holistic interpretation of the researchers was verified using a quantitative approach. Six judges evaluated the interpretability of positive and negative reviews, and their assessments were compared for inter-rater reliability. The mean interpretability scores for positive and negative reviews were compared.
5. The next step was the distributional analysis of topic loadings in the positive and negative reviews and its visualization. The behavior

patterns of probabilities of a particular review were loaded on a particular topic and compared for positive and negative reviews.

6. Finally, to place more confidence in the findings, the same analytical approach was applied to the reviews of Red Square in Moscow, Russia and Chichen Itza in Mexico. To save space, only the most immediate results for two additional data sets are included in the body of the manuscript and the rest is presented in Appendix. In addition, we present only analysis of the English language reviews.

3.1. Terracotta Army data

We collected 14,273 visitor reviews from 2011 to 2019 from the TripAdvisor site of the Terracotta Army museum. The collected reviews were referenced to the reviewers' home countries in the following way. First, the text describing the geographic location of the reviewer was resolved into latitude and longitude using the Google Geolocational API. Next, the geographical coordinates were reverse-geolocated into countries using OpenCage API (OpenCage.com). In total, the location was found for 11,242 reviews (79% of the total) left by visitors from 129 countries. Most of the reviewers came from the USA (19.7%), followed by Mainland China excluding Hong Kong and Taiwan (15.0%) and the UK (11.6%). For further processing, two subsets were generated: "En" subset (5859 reviews) representing English language reviews coming from visitors of five industrialized and culturally close countries (USA, UK, Australia, Canada, and New Zealand) and "Zh" subset (1792 reviews) representing simplified Chinese reviews by visitors from the hosting nation (Mainland China; zh is the ISO code for Chinese). Each review comes with a star-rating of 1–5 expressing the visitor's evaluation of the attraction, with a score of 5 being the most favorable. Table 1 shows star-rating distributions of the collected En and Zh reviews.

We interpreted 1-3-star reviews as negative and 4-5-star reviews as positive. The rationale for our decision is two-pronged. First, there is an abundant evidence that reviews in any star rating category contain a mix of positive and negative statements, so that treating the "mid-rating" reviews as neutral is unsupported (Fong, Lei, & Law, 2017). Second, market research shows that it is too "generous" to count only one- and two-star reviews as failure on the part of the business because only 13% (Marchant, 2015) to 18% of customers (Murphy, 2019) would consider using a business with the mean star rating below three. Moreover, only slightly over half of the customers would even consider a business with a star rating below four (Marchant, 2015; Murphy, 2019), justifying that a three-star business is still close to the failing ones for almost a half of potential customers.

The English and Chinese pools of reviews were then divided into subsamples of 4-5-star reviews and 1-3-star reviews, hence creating four subsamples. It can be seen from Table 1 that there is a dramatic difference in the numbers of positive and negative reviews in both En and Zh samples. We also compared the length of the reviews. Negative reviews had larger means in both En and Zh samples across all star-rating groups (Table 1). For the original Chinese logograms, the length of the negative reviews was also slightly larger. The same pattern regarding the number of positive and negative reviews and their length was observed in the Red Square and Chichen Itza data (Table A1 in the Appendix).

3.2. Unsupervised data mining algorithm: Latent Dirichlet Allocation (LDA)

The main topics discussed by the visitors to the Terracotta Army site were extracted in the following way. First, the Zh dataset was translated from Simplified Chinese to English using Google Translate. Google Translate has been previously employed for TripAdvisor review analyses (e.g., Song, Kawamura, Uchida, & Saito, 2019). While the purposeful translation quality analysis for tourism reviews has not been studied, in other domains like school essays (Valijärvi & Tarsoly, 2019) and the UN documents (Windsor, Cupit, & Windsor, 2019) the outcomes were reported as satisfactory.

Second, all collected reviews were pre-processed using the standard steps for computer-assisted content analysis (Heydt, 2018; Manning, Raghavan, & Schütze, 2008):

1. Tokenization with removal of short (below 4 symbols) and long (above 25 symbols) tokens;
2. Filtering English stopwords (from Page Analyzer, https://www.ranks.nl/stopwords);
3. Filtering tokens by parts of speech, retaining nouns and adjectives;
4. Stemming (reducing inflected words to their word stems using Porter stemmer);
5. Transformation of words to lower case;
6. Transformation of words with variant spellings (e.g., terracotta and terra cotta);
7. Filtering the most frequent and infrequent words. Words encountered in over 70% and lesser than 1.5% of documents are excluded as not discriminating among the topics.

Third, the main topics were extracted with LDA as implemented in the MALLET package (http://mallet.cs.umass.edu) with optimized topic density/words density parameters. Two indices were used to compare performance of LDA topic models: perplexity and coherence. Perplexity measures how well the word distribution predicted by the LDA model matches the actual word distribution. That is, perplexity measures how well the outcomes of the theoretical model built on the underlying LDA assumptions matches the observations. The coherence index measures the semantic similarity between the high-ranking words in a topic. Specifically, the coherence evaluates the frequencies with which the high-ranking and lower-ranking words composing the same topic tend to appear together in documents.

The LDA algorithms optimize distribution of the words in topics and topics in documents given a pre-set number of topics K; hence the value of K affects LDA results. The automated methods that allow for determining the "best" K based on perplexity minimization are known to return very large K values, roughly equal to 0.05 of the number of

Table 1
Distribution of star-rating and length of reviews for Terracotta Army data: English speaking (En) countries and Mainland China (Zh).

Terracotta Army	Reviews per sub-sample				Words per review			
Review	En		Zh		En	En	Zh[a]	Zh[a]
Star-rating	N	%	N	%	Mean	Median	Mean	Median
1	11	0.2	5	0.3	186.4	111	84	69
2	38	0.6	13	0.7	148.6	105	109	95
3	149	2.5	156	8.7	120.2	92	101	77.5
4	785	13.4	605	33.8	103.4	76	92	76
5	4876	83.2	1013	56.5	81.2	56	89	72
Negative reviews: star-rating 1, 2, and 3	198	3.3	174	9.7	129.3	94.5	101.5	78.5
Positive reviews: star-rating 4 and 5	5661	96.6	1618	90.3	84.3	59	90	74
Total	**5859**	**99.9**	**1792**	**100.0**	**85.8**	**60**	**91**	**74**

[a] Number of Chinese logograms.

documents, frequently resulting in thousands of poorly interpretable topics (Chang, Gerrish, Wang, Boyd-Graber, & Blei, 2009). Meanwhile, to be practical, topic modeling needs to return topics that are interpretable (understood) by humans. Hence, the majority of practical applications employ a manual method to determine the value of K that returns a small number of topics that are highly interpretable (Newman, Lau, et al., 2010). In our research we used a combination of both approaches known as the "elbow method." As K increases, topic coherence tends to improve; this improvement is fast when K is small and slow when K is large. The elbow method evaluates K as the number of topics approximately corresponding to the "elbow" on the coherence (K) function plot (Fig. 1). Mathematically this elbow corresponds to the maximum of the second derivative of the function. The precise value of K is then determined based on an "interpretability" of derived topics and their "optimal" separation, which means that topics are neither widely agglomerated nor are they redundant. This procedure is inherently qualitative and "holistic," and its purpose is to address point #4 in the guidelines by Tang et al. (2014) that extracting overly large number of topics should be avoided.

Following this procedure, we first determined the optimal value of K to be between 10 and 15 from the "elbow" criterion analysis, then qualitatively interpreted results of LDA models with K = 10 … 15, and finally selected the K value corresponding to the best topic interpretability and separation. Hence, the optimal K value was set at K = 14 for En reviews and K = 11 for Zh reviews.

4. Results

4.1. Terracotta Army reviews: topic extraction

Applying the method and parametric criteria described in the previous section, results presented in Table 2 (En dataset) and Table 3 (Zh dataset) were obtained.

4.2. Terracotta Army reviews: interpretation of LDA solution

Tables 2 and 3 show the topics identified from the En and Zh datasets, respectively. While the topics arising from the positive (4–5 star) reviews are easily interpretable based on words comprising those topics, the majority of topics arising from negative (1–3 stars) reviews are difficult to interpret based on the words alone. Manual analysis of the negative reviews which load highly on those topics found multiple complaints about a large variety of service failures. In the En, but not the Zh dataset, the most commonly shared complaint was the crowdedness

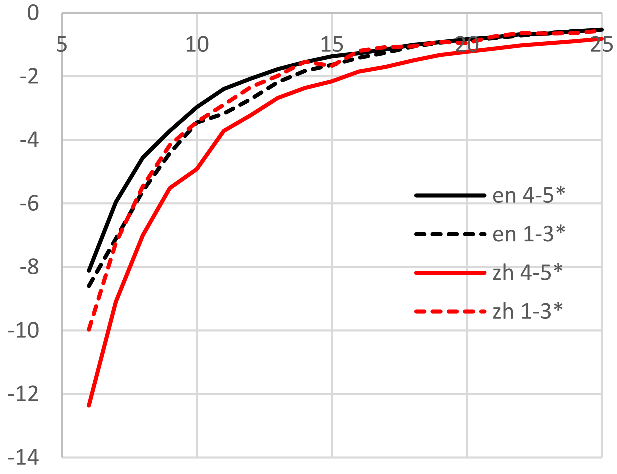

Fig. 1. The change in topic coherence as a function of the number of topics for English (En) and Chinese (Zh) datasets; 1–3 and 4–5 star reviews are shown separately.

Table 2
Topics extracted from En sample. Ten tokens most closely associated with the topic are presented.

Topics: 4–5 star reviews	Tokens
Culture, 8th Wonder of the World	site world china wonder history archeology great ancient visitor culture
Farmer signs the book, gifts	farmer book picture warrior shop sign site gift terracotta souvenir
The place is worth to visit	place visit amazing worth time history picture museum site guide great
Travel logistics	city Beijing wall Xian time hour china trip airport hotel train Shanghai
Excellent private tour	guide tour museum informative private group time history knowledge excel
Amazing original excavation	warrior work pit archeology piece site restore excavate horse amaze original
Area logistics, shops and food	shop walk museum souvenir restaurant food price area good park gift
Clay army description	emperor soldier terracotta warrior hors army tomb chariot life china face
Impressive museum	warrior museum site terracotta visit area pit impressive hour attract main
Main excavated pit of museum	warrior hall exhibit picture pit chariot display walk view good glass main
Busy place, crowds esp. mornings	crowd people time busy place holiday picture good site morning front
Getting there locally	station ticket train yuan entrance bus hour taxi stop cost museum guide
Amazing life-size warrior army	warrior size amazing detail terracotta scale mind sheer incredible life army
Great experience worth a travel	warrior china terracotta amazing trip great experience visit worth highlight
Topics: 1–3 star reviews	**Tokens**
Disorganized place	clear lack background organize admission work constant empire commission
Do not trust local bus operators	operator problem terracotta cash town bus larger green term resident
Hope for future improvement	nice expect complex quality hangar future opinion actual type unexcavated
Multiple criticisms[a]	visitor disappoint moment camera vendor fact actual fake room massive
Fantastic place, but[b] …	figurine queue hawker hanger extra Quin theatre rude box pottery fantastic
Multiple criticisms [c]	guide museum tour terracotta Xian information shop army driver soldier
How to get there. Overrated.	walk ticket station train yuan cost park number sign price stop wall
Multiple criticisms[d]	hall color weapon mausoleum task film dynasty example shot cool trap
No discount for foreign students	student charge government store price country factory replica pick amount
Multiple criticisms[e]	preserve light famous control commerce able safety extreme advance pity
Multiple criticisms[f]	warrior site china people terracotta picture visit tourist crowd place pit time
Multiple criticisms[g]	cart ride golf highlight video seller floor movie course corridor plan extra
Multiple criticisms[h]	wall world total book guess food easy Shanghai other wonder skeptic
Local businesses[i]	warehouse vast copies giant awesome event card heritage hold bronze

[a] Sketchy story, likely fake, waste of time, noise, crowds.

[b] Crowded, busy, rude people, pushing, lack of organization, no toilet paper, fake, etc.

[c] Most information in Chinese, too much information from guides, terrible guides, overrated, long drive.

[d] Probably fake, crowded, poor documentary movie, best colored figurines are missing, etc.

[e] Uncontrolled local commerce, seedy street from the station, bad display inside, crowds, rude tourists.

[f] Distance to figurines, photos misrepresent the site (figurines not colored, site size overstated etc.), crowds, local tourists cutting into lines, etc.

[g] Poor management: gold cart rides to museum, information video, missing plans, dump, hot, crowds.

[h] Crowds, pickpockets, little to see, broken movie theater, local commerce, fake. Tiring trip from city.

[i] Rude and overpriced local businesses, real/fake farmer who discovered the site selling overpriced autographed books, expectations to pay for everything. Many of displayed artefacts are copies.

Table 3
Topics extracted from Zh sample. Ten tokens most closely associated with the topic are presented.

Topics: 4–5 star reviews	Tokens
Description of the figurines	figurine color unearth face thousand expressive look amazing life
Guided tour: Pros and cons	tour guide people good time price ticket group student look money talk
Museum and excavation pits	museum hall exhibit archeology hole amazing walk site pit travel antique
Tourism environment and services	good area attract beautiful scenic tourism environment place spot regret
Pride, 8th wonder of the world	people ancient china world wonder great history wisdom culture eighth
Emperor's tomb, historic place	history tomb soldier sense emperor general imperial front king face
Descriptions of terracotta warriors	terracotta warrior horse spectacular feel time look museum figurine
Visit w/friends, children; foreigners	time foreign summer school travel love child friend door experience
Shock and emotions	feel people history ancient dynasty shock terracotta place momentum kind
How to get there	station railway train ticket hour yuan convenient line direct area morning
Worth a visit	visit worth attract history tourist feel good look place foreign interest
Topics: 1–3 star reviews	**Tokens**
Multiple criticisms[a]	admire scenery play dirt distance wave ancient native book similar
Expensive, nothing special[b]	price terracotta face loess visitor help communist fine pile okay work
Multiple criticisms[c]	scenic care capital driver phenomenon travel carriage copper period
Great place but something is not right	history people china ancient great place culture worth world foreign site
Multiple criticisms[d]	rain weapon guide hand vehicle hour result Xiang went environment
Spectacular, but hard to see, costly	terracotta warrior horse look feel time spectacular figurine people attract
Problems with accessibility	station tourist convenient individual train store railway sale tour dollar
Spectacular, but undeveloped	exhibit technology undeveloped mausoleum spectacular spring vote area
Management problems	lack management effect protect visitor majestic downtown software
Tour dissatisfaction	good guide tour visit regret people feel ticket expensive place museum
Tickets: lines, price, scalpers	ticket time office free dollar scalper local hole wood obvious opportunity

[a] The documents loading to this topic while expressing some admiration also contain multiple criticism points, including: just ancillary services, many wild tour guides, soliciting, toilets are flawed, during the New Year there are many people, but only half of the site open, boring, crowded, violent people, place probably fake, just a dirt pit, and others.

[b] An example of a loading review: the place is ok, but I just go for a local craft shop.

[c] Touching on all issues but in various ways: wanted to see copper carriage but was pushed away.

[d] Expensive tickets, rain, site destroyed by Xiang [Yu rebellion] - all sorts of complaints.

of the site. Indeed, "crowd" was the most frequent negative word in En 1 – 3 star reviews and ranked 13th in the overall word frequency distribution; the same word was also used frequently in the positive reviews and ranked 20th.

Connected to the crowdedness issue, the En dataset also contained multiple complaints about tourist behavior: pushing, rudeness, cutting into the lines, and so on. The majority of other negative review topics were similar to the En and Zh datasets. The most commonly shared complaints involved local entrepreneurs, including soliciting, pushing sales, dishonesty, and a seedy business street leading from the station to the museum. Interestingly, many reviewers mentioned a local entrepreneur positioning himself as a farmer who discovered the site in 1974 and who was selling autographed books; some reviewers seemed to believe the farmer was genuine, while many others commented he was an imposter. Finally, English and Chinese visitors alike complained about prices; in addition, visitors noticed that while the Chinese student tickets were half-priced, foreign students were not allowed this discount.

Other complaint topics, however, were scattered. Those negative reviews were related to prices for various services and management problems (tour guide confusion, poor quality software, dirty restaurants, partial closings during the peak seasons, ticket lines, and ticket scalpers at gates, among many other complaints). Many reviewers disputed the authenticity of the site, ranging from its commercialization ("Disney attraction rather than an archeological treasure") to its genuineness as a whole ("a ploy for Communist China to gain tourism"). Some negative reviewers from both data sets believed that the museum artefacts were either replicas or entirely forged. Another criticism was connected with interpretations of the site's history by the museum guides being misaligned with those found in history books. Complaints about traveling to the museum site were also diverse and included highway traffic, bus drivers miscommunicating the route to get more passengers, frightening taxi rides, and a ride from the train station in a golf cart. Finally, many reviews expressed general disappointment: boring, just a dirty pit, broken pottery, missing figurines, and others (see comments to Tables 2 and 3).

Meanwhile, the majority of the dissatisfaction topics in the 1–3 star reviews were left unidentified in the LDA analysis and required additional consulting with the original reviews (Tables 2 and 3). Note that the objective measures of topic coherence (Fig. 1) did not show significant differences in the coherence of the topics arising from positive and negative reviews, suggesting that the LDA topical modeling was similarly successful for all four datasets. To confirm that this effect is due to the nature of negative reviews, as compared to the positive reviews, and not to sample size differences, we randomly selected two samples from 4 to 5 star En and Zh datasets with sizes matching the 1–3 star samples. We ran LDA analyses on those samples and found the topics arising from those samples interpretable and similar to those in the full datasets.

4.3. Validating study results with Red Square, Russia, and the Chichen Itza, Mexico, reviews

Analyses of two additional attractions (Red Square in Moscow, Russia and Chichen Itza in Mexico) were performed to confirm the nature and interpretability of negative reviews. We obtained the reviews in the English language for these two sites: 7700 for Chichen Itza from visitors from the USA, Canada, Australia, New Zealand, and UK and 3529 for Red Square from visitors from the same countries plus the EU. The LDA solution presented in Appendix Table A2 shows the same pattern of a significant number of uninterpretable topics in both review pools. Only two topics identified by LDA in the negative reviews of the Chichen Itza site were clearly negative: noise and a large number of local vendors. The rest of the topics were either neutral, positive, or unidentifiable. A qualitative analysis of the reviews, however, discovered multiple other problems noticed by the visitors: overcrowding, very hot, long lines, fences, misbehaving tourists, bored tourist groups, high prices, noise, conflicting information, questionable restoration and many others.

Even a larger number of unidentifiable topics were found among the Red Square negative reviews: 9 out of 14. Among those topics that were successfully interpreted, only one was clearly negative: problems with food and dining options. Similar to the Chichen Itza site, a qualitative analysis discovered multiple other problems experienced by the visitors: too many tourists, uneven hazardous paving, confusing church interior, congested Lenin mausoleum, fences, construction work, rude staff, odd closing times, strange security protocols, high commercialization, boring, just not interesting (*"It's a square"*), and many more. These experiences were not reflected by LDA in separate themes being either unique or expressed in vastly different language such as "*fencing of parts of the square for construction or various events."*

4.4. Quantitative assessment of interpretability by human raters

To validate interpretability of the obtained topics as holistically assessed by researchers, topic interpretability was independently analyzed by seven human raters following methodology offered by Newman, Lau, Grieser, and Baldwin (2010). Each rater analyzed all 106 topics discovered by LDA: 50 for Terracotta Army, 28 for Chichen Itza, and 28 for Red Square data sets. The raters knew that the research purpose was to investigate human ability to interpret computer-generated LDA review topics. Each reviewer was presented with a list of words comprising generated topics but saw neither the reviews loading on those topics nor whether the topic came from a positive or negative set of reviews. For each location and language, topic order was randomized so that topics from the positive or negative reviews were encountered by reviewers in no particular order. Following Newman, Lau, et al. (2010), the raters were instructed to rate topic interpretability as 1 (poor), 2 (medium), or 3 (good). For a complete questionnaire and rating criteria, see Appendix Box 1.

The summarized results averaged across all reviewers are shown in Table 4. One-tail *t*-test was conducted to test hypothesis that negative reviews had lower interpretability score as compared to positive reviews. The results from the positive and negative reviews indicate that the difference between mean positive M^+ and negative M^- interpretability scores (Table 4) vary between 0.37 and 1.06 points and is statistically significant at 0.01 level. The smallest difference was indicated for the Chichen Itza site: $M^+ - M^- = 0.37$, $t(6) = 4.3$, $p = .006$. One reason for that may be that 4 out of 7 raters indicated that they were not aware of this site. Note, however, that the negative reviews for this site had several topics with positive content (great historical site but with problems of local vendors) with large number of documents loading on those which we discussed: see Discussion and Fig. 2 in the next section. For all other sites the difference between interpretability of positive and negative reviews of approximately 1 point was statistically significant at a 0.001 level (Table 4).

Individually, all reviewers evaluated interpretability of negative reviews worse than those of positive reviews for all sites (see Appendix Table A3 for individual scores). In addition, the percentage of reviews with poor interpretability was always larger for negative reviews and the percentage of reviews with good interpretability was always larger for positive reviews (Table 4). Overall, the results of content analysis validated the previously stated conclusion that negative reviews have consistently poorer interpretability of topics derived with LDA analysis.

4.5. Different nature of positive and negative reviews: distribution of topic loadings

In the LDA analysis, representation of a topic in a specific document is expressed as topic loading. The documents highly representative of a specific topic would have high loading on this topic and low loading on other topics. If a large number of visitors mention the same issue (positive or negative), ideally, the LDA algorithm will identify it as a topic, and there will be a large number of reviews with high loadings on that topic. When a topic is rarely mentioned in visitors' reviews and is, in fact, a statistical artefact, then there will be just a few reviews that load high on that topic and many reviews with low loadings. The comparison of loading distributions for identified topics in the negative and positive reviews allows for additional insights into the nature of those reviews as well as the applicability of the LDA method for either type of reviews.

Fig. 2 presents patterns of topic loadings in the collected documents for all three attractions. Positive reviews are similar for the three sites; universally, the red lines start high and gradually decline. Topic loadings in negative reviews demonstrate a more complicated behavior pattern. Loadings for a majority of the topics drop quickly – observe the blue lines close to the horizontal axis. A few negative topics, however, exhibit a pattern similar to that of red lines. This behavioral pattern for topics from negative reviews is similar for Terracotta (En and Zh), Red Square, and Chichen Itza datasets.

5. Discussion

The study found support for its original proposition that the nature of positive and negative tourist reviews obtained from TripAdvisor (and, supposedly, similar review platforms) is different enough to call for increased awareness in application of unsupervised data mining algorithms like LDA for topic extraction. The results of LDA analysis show that 4–5 star topics in general are easily interpreted even without referring to the original reviews loading on those topics. For the Terracotta Army site, for example, visitors comment on the historical significance of the place, a feeling of amazement observing the huge clay army, logistics of the travel, and local services. In addition, many Chinese visitors comment on the feelings of pride for their history and people. The panel of reviewers found interpretability of 67% of topics (on average) generated from 4 to 5 star reviews to be "good" and only 6% "poor". Conversely, 42% of negative reviews on average were poorly interpretable and only 22% highly interpretable.

The percentage split between poorly and well-interpretable topics in negative reviews could have been even greater, if not for the definition of a negative review. While some researchers consider only 1 and 2 star reviews as negative (e.g., Cenni & Goethals, 2017), others, including researchers in this study, add 3-star reviews as well (Chang, Ku, & Chen, 2019). We maintain that the three-star reviews still contain elements of a failed service and, hence, should be investigated by business management in addition to the 1- and 2-star ones. What is also important, 3-star reviews are more numerous (Tables 1 and A1), which improves

Table 4
Results of topic interpretability study, averaged over reviewers. The Interpretability columns show percentage of reviews with interpretability scored as poor, medium, and good, and the Mean column represents the mean interpretability rating averaged over all topics and reviewers. Last two columns represent results of a two-sample paired *t*-test (N = 7).

Dataset	subset	N	Interpretability			Mean	Diff.	Std.	t(6)	p (one-tail)
			Poor	Med.	Good					
Chichen Itza	negative	14	26%	38%	37%	2.1	0.37	0.22	4.3	0.002
Chichen Itza	positive	14	11%	30%	59%	2.5				
Red Square	negative	14	49%	35%	16%	1.7	1.06	0.29	9.6	<0.001
Red Square	positive	14	3%	20%	77%	2.7				
Terracotta, Zh	negative	11	42%	39%	19%	1.8	0.81	0.13	16.0	<0.001
Terracotta, Zh	positive	11	5%	31%	64%	2.6				
Terracotta, En	negative	14	51%	35%	14%	1.6	1.04	0.14	19.4	<0.001
Terracotta, En	positive	14	3%	27%	70%	2.7				

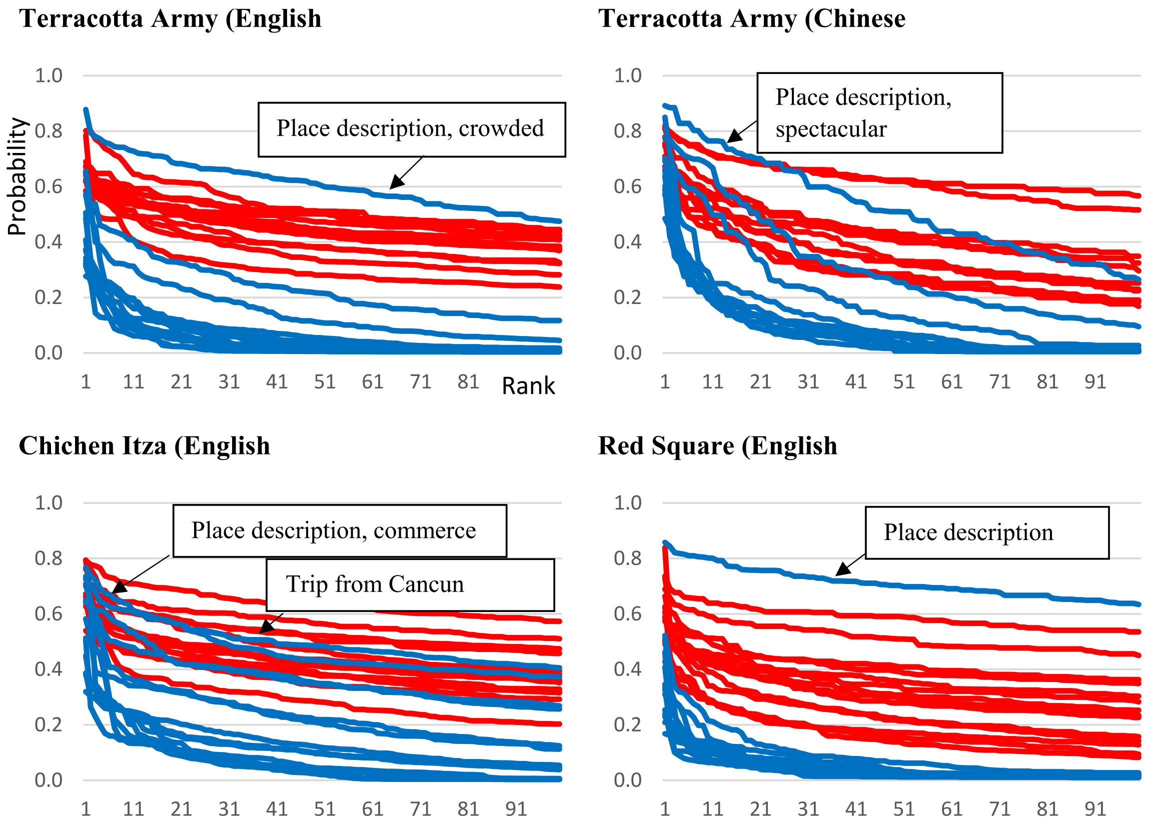

Fig. 2. Distribution of topic loadings for negative (blue) and positive (red) reviews. Each curve represents a single topic. The horizontal axis represents the top 100 documents sorted in order of decreasing probability of representing a particular topic and the vertical axis represents the probability. Notice that almost all negative reviews have very low probability to belong to any particular topic except for a few common topics (see the text boxes). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

LDA analysis from the algorithmic standpoint. Whatever the definition, we observed some presence of positive topics even in 1-star reviews, similarly to Dinçer and Alrawadieh (2017) and Vásquez (2011). This observation is also in agreement with Fong et al. (2017) study that found dual-valence (that is, containing both positive and negative sentiment) reviews existing in all five hotel rating categories on TripAdvisor. We found that the few positive topics present in negative reviews were shared among multiple documents (top blue curves on Fig. 2). For example, the most common topic in Chinese sample of the Terracotta Army set of negative reviews is the description of the spectacular statures (*terracotta warrior horse look feel time spectacular figurine people attraction*) (Table 3). The researchers see the presence of positive topics in negative reviews as reason why on Fig. 2 one or a few negative (blue-color) lines are above all others. In contrast, negative topics anchored by a few tokens often did not check well across original reviews. For example, the token *tour guide* was associated in the tourist reviews with insufficient information, not enough tour guides, bad translation, wrong information, logistical problems, etc. (Table 3). Loadings for such topics dropped quickly in the original documents, which is reflected in blue lines close to the abscissa on Fig. 2. These results clearly indicate that negative reviews, in general, can be problematic with respect to topic interpretability.

Overall, we obtained strong support for our initial proposition that the results of the LDA analysis of data sets with low customer satisfaction are significantly less interpretable as compared with the analysis of the positive reviews. One reason is a smaller number of negative reviews as compared to positive reviews; the smaller population size complicates analysis of the recurrent patterns. This seems a common problem. In analyses of travel reviews, LDA has been applied to datasets of radically different sizes, ranging from as few as 50 (Putri & Kusumaningrum, 2017) to over 250,000 (Guo et al., 2017) reviews. The latter article regressed the review star rating on 19 review topics extracted with LDA for the purpose of finding the most important dimensions of tourist satisfaction. The dataset size (266,544 reviews of 25,670 hotels located in 16 countries) clearly fits within the guidelines suggested in Guo et al.'s (2017) research, which operated with a corpora consisting of only a few thousand documents. Restricting LDA analysis to large cumulative datasets is, however, impractical in some contexts. A single popular destination typically contributes several hundred to several thousand reviews annually as evidenced from TripAdvisor pages. Dividing the collected dataset into several groups, e.g., by the expressed sentiment, may further reduce subset sizes. In their analysis of Indonesia tourist reviews, Putri and Kusumaningrum (2017) collected 100 reviews, which were further divided into positive and negative review classes of 50 reviews each.

Yet another reason, we argue, is the Anna Karenina principle, that is, a much higher diversity of dissatisfaction issues in negative reviews. Together, these two features of negative reviews result in topics with "not so good" separation and, consequently, low interpretability. The researchers obtained the same result for three culturally different attractions and two languages (for the Terracotta Army set). More research is needed to test the proposition in various contexts and for various sample sizes. If, however, the proposition is well established across multiple settings, it would mean that the Anna Karenina principle limits application of the automated topic modeling when the nature of data, particularly, large between – document variability, makes the unsupervised classification methods not well suited for their processing. Specifically, the analysis of the main topics of customer dissatisfaction requires collection of very large datasets where the sheer volume of reviews would warrant keeping a much greater number of topics arising from the LDA analysis. When large datasets are unavailable and their collection is not justifiable, using the qualitative research methods to find the main categories of customers' complaints followed by supervised classification is advisable.

5.1. Managerial implications

Destination image building is based on its most distinguishing features and/or resources, thus, differentiating a destination from its competitors (Beerli & Martin, 2004; Hosany, Ekinci, & Uysal, 2006; Moreira & Iao, 2014). In this study of the Terracotta Army attraction, positive reviews converge on "the 8th wonder of the world," "life-size warrior army," "the legendary digging by farmers," "amazing excavation," and other history-related aspects of the attraction. Obviously, these features of the Terracotta Army brand should be maintained and highlighted. Contrary to the highly concentrated content of positive reviews, more dispersed, topic-wise, negative reviews present problems for the customer service departments and the marketers. Although every complaint may reflect a real problem faced by its reviewers during the visit, it is impossible for attraction managers to respond or deal with every complaint separately. Thus, it is critical to identify types of complaints and reasons why visitors feel unsatisfied and then propose specific recovery strategies and approaches that can improve tourists' experiences and, more importantly, maintain the Terracotta Army brand. Thus, the study offers attraction managers a few insights based on our reading and interpretation of the data.

First, the study strongly suggests that application of unsupervised topic mining algorithms to negative reviews may be problematic, depending on attraction specifics; therefore, the results should be treated with caution. Manual verification of results is advised, especially in a situation when tokens pertaining to a particular topic are difficult to interpret. To combat a situation when core tokens in the topic are related to different issues of dissatisfaction, a dictionary-based approach is recommended. Managers may create a list of issues (and related words) they want to track, and the algorithm would calculate how many reviews from the negative pool mention a particular issue. The dynamics of those issues – and the effectiveness of a managerial action – can be tracked. This approach is quite flexible, as it can be instigated on a regular basis to monitor crowdedness, transportation problems, commercialization, etc. and detect public sentiment and improvement. The other principally different tactic to tackle the negative reviews, would be to employ a supervised algorithm. These algorithms can be trained on similar data, say, from previous periods, and used for classification of negative reviews from the current period in topical categories. This approach, however, needs investigation and validation.

Second, we recommend attraction managers focus on systemic problems in complaints that are likely to reflect faults in business model or operation philosophy of a tourist attraction company. Our study shows that a number of complaints presented in Tables 2 and 3, are problems that are reflective of a business model of profit-oriented utilization of heritage in the Terracotta Army museum and reflect an unbalanced relationship between heritage conservation and tourism commercialization. Previous studies have shown that over-commercialization affects tourists' perceptions of heritage authenticity and their satisfaction (Breathnach, 2006; Hughes & Carlsen, 2010). Note that people tend to be more dissatisfied when they deem the business in control of the point of dissatisfaction (Jiang, Gretzel, & Law, 2010, pp. 297–308). Therefore, attraction managers might consider giving up some profits in pursuit of a heritage-centered business model. If this is the case, the scope of tourism businesses and the number of visitors to the Terracotta Army might be limited.

Third, it is necessary to analyze the time attribute of complaints, differentiating between temporary issues and permanent, persistent problems. When an issue is mentioned repeatedly, it poses a threat to the reputation and image of an attraction. In the case of the Terracotta Army, complaints about tour guides, soliciting local businesses, and ticketing have appeared for a prolonged period of time. Although service recovery efforts are, undoubtedly, vital in the hospitality industry, the adverse impact caused by repeated service failure can hardly be mitigated by recovery measures, for example by being courteous (Liao, 2007).

It should also be noted that the dispersion of complaints has a spatial dimension as well. For the effectiveness of governance, verification of the geographical location where complaints have a tendency to occur is conductive to the implementation of different countermeasures, with a possible involvement of both the attraction management and municipal administration. Recall, for example, the persistent complaint about less than adequate transportation from Xi'an to the Museum or complaints related to fake tour guides wandering outside of the train station who mislead tourists to get them to buy tickets, but the tickets were actually just for other low-level attractions along the road, excluding the Terracotta Army. These types of complaints require a coordinated effort of various tourist agents and branches of government to stop cheating behavior and to maintain the image of the top national attraction in China. We propose it as a topic for future research.

Credit author statement

Andrei P. Kirilenko: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Supervision; Visualization; Writing - original draft; Writing - review & editing. Svetlana O. Stepchenkova: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Roles/Writing - original draft; Writing - review & editing. Xiangyi Dai: Data curation; Investigation; Validation; Writing - review & editing.

Impact statement

Online travel reviews are increasingly being used by scholars and industry to identify and address issues related to travelers' satisfaction. The challenges of analyzing large quantities of social media data make automated identification of the main topics of online reviews desirable; however, the methodological guidelines have not been developed. This study establishes a major restriction related to automated extraction of factors of tourist dissatisfaction from online reviews. We demonstrate generalizability of these restrictions on three geographically, linguistically, and culturally diverse tourist attraction and connect them to the overarching Anna Karenina Principle governing the factors of success and failure. The study strongly suggests that application of unsupervised topic mining algorithms to negative reviews may be problematic, depends on attraction specifics, and the results should be treated with caution.

Declaration of competing interest

None.

Acknowledgments

The authors are thankful to the following University of Florida graduate students who helped with interpretability analysis and validation of the model: Shihan (David) Ma, Lijuan Su, Humeyra Dogru, Hyejin Park, Jing Yang, and Yuting An.

Appendix A

Table A1
Distribution of star-rating and length of English language reviews for Chichen Itza (USA, Canada, Australia, New Zealand, and UK) and Red Square (USA, Canada, Australia, New Zealand, UK, and EU) data.

	Reviews per sub-sample				Words per review			
	Chichen Itza		Red Square		Chichen Itza		Red Square	
Star-rating	N	%	N	%	Mean	Median	Mean	Median
1*	58	0.8	3	0	157.1	125	56.7	31
2*	156	2	23	0.4	178.1	132	73.4	52
3*	433	5.6	161	2.6	155.8	122	55.0	43
4*	1788	23.2	788	12.9	146.8	108	54.7	43
5*	5265	68.4	2554	41.7	121.5	86	52.6	39
Negative reviews (1*-3*)	647	8.4	298	4.9	161.3	126	57.3	44
Positive reviews (4*-5*)	7053	91.6	5827	95.1	127.9	91	53.1	40
Total	**7700**	**100**	**6125**	**100**	**130.7**	**94**	**53.3**	**40**

Table A2
Negative review themes, Chichen Itza (Mexico) and the Red Square (Moscow, Russia) sites. The tokens represent the highest ranking words.

Chichen Itza	Tokens
Cenote swimming*	cenote lunch stop time hour buffet town swim food water
Uninterpretable	Castillo Maya warrior power harass Toltec apple date platform equinox
Amazing but ruined by vendors*	ruin vendor place experience people history sell amazing great tourist
Getting the ticket*	card credit ticket cash peso change receipt charge phone line
Tours (Amigo, Flavio, Viator)	amigo transport talk check Flavio shot beware specific tequila Viator
Amazing but ruined by vendors*	site vendor Chichen visit tourist experience world visitor area hundred
Uninterpretable	picture Mayan toilet info guide water price haggle work video
Trip from Cancun*	tour guide hour Chichen trip time Mayan good hotel Cancun
Noise	jaguar noise sale table pitch whistle Mexican loud sound constant
Entrance	entrance ticket park peso road line toll sign cost govern
Cruise tours	cruise seat ferry ship Cozumel line Carnival passenger excursion operator
Uninterpretable	light night sound waste Spanish dark even cost idea joke
Site description*	ruin Chichen visit pyramid time Mayan history site structure people
Uninterpretable	rope Palenque hire museum climb access touch tour day picture
Red Square	**Tokens**
Uninterpretable	capital Russian corner doubt icecream [dome] group day rule formal private
Uninterpretable	color backdrop shoulder reason train expensive destination deal land funny
Site description*	square visit Moscow place Basil Kremlin cathedral time tourist Lenin
Uninterpretable	seat room money stand water firework international unlucky symbol thank
Uninterpretable	side wall crowd previous iron leader camera ceremony center build
Uninterpretable	side look scaffold worker complete minute competition rest trouble
Uninterpretable	mini kind chance walk occasion discrete bozo sell nuclear rain
Uninterpretable	tower former state want future help fence bleacher quick mausoleum
Kremlin	rate Kremlin tank wish photography seat classic cross proportion
Uninterpretable	shop great line church music number language badge cleaner favorite
Uninterpretable	empty saint tomb time heavy lens hand avoid impact mega
Eating problem	waiter expensive third time name floor juice reason picture impossible
Military parade	parade make space Russian tank sight smaller diameter television impress
Soviet time reminiscence	cool missile local amount ready Soviet roll clear various

[*] High loading topics.

Table A3
Mean scores and percentage of reviews with poor, medium, and good interpretability returned by individual raters.

Rater	Site	Subset	N	Mean	Interpretability		
					Poor	Medium	Good
1	Chichen Itza	negative	14	1.71	43%	43%	14%
1	Chichen Itza	positive	14	2.36	21%	21%	57%
1	Red Square	negative	14	1.21	79%	21%	0%
1	Red Square	positive	14	2.79	0%	21%	79%
1	Terracotta, zh	negative	11	1.45	64%	27%	9%
1	Terracotta, zh	positive	11	2.36	18%	27%	55%
1	Terracotta, en	negative	14	1.64	50%	36%	14%
1	Terracotta, en	positive	14	2.64	7%	21%	71%
2	Chichen Itza	negative	14	2.36	21%	21%	57%
2	Chichen Itza	positive	14	2.86	0%	14%	86%
2	Red Square	negative	14	1.93	21%	64%	14%
2	Red Square	positive	14	2.93	0%	7%	93%
2	Terracotta, zh	negative	11	1.91	36%	36%	27%
2	Terracotta, zh	positive	11	2.82	0%	18%	82%
2	Terracotta, en	negative	14	2.00	14%	71%	14%
2	Terracotta, en	positive	14	2.86	0%	14%	86%

(*continued on next page*)

Table A3 (*continued*)

Rater	Site	Subset	N	Mean	Interpretability		
					Poor	Medium	Good
3	Chichen Itza	negative	14	1.93	29%	50%	21%
3	Chichen Itza	positive	14	2.07	21%	50%	29%
3	Red Square	negative	14	1.36	71%	21%	7%
3	Red Square	positive	14	2.50	7%	36%	57%
3	Terracotta, zh	negative	11	1.45	55%	45%	0%
3	Terracotta, zh	positive	11	2.36	0%	64%	36%
3	Terracotta, en	negative	14	1.21	79%	21%	0%
3	Terracotta, en	positive	14	2.50	0%	50%	50%
4	Chichen Itza	negative	14	1.57	50%	43%	7%
4	Chichen Itza	positive	14	1.79	36%	50%	14%
4	Red Square	negative	14	1.07	93%	7%	0%
4	Red Square	positive	14	2.36	14%	36%	50%
4	Terracotta, zh	negative	11	1.36	73%	18%	9%
4	Terracotta, zh	positive	11	2.18	18%	45%	36%
4	Terracotta, en	negative	14	1.21	79%	21%	0%
4	Terracotta, en	positive	14	2.14	14%	57%	29%
5	Chichen Itza	negative	14	2.50	7%	36%	57%
5	Chichen Itza	positive	14	2.79	0%	21%	79%
5	Red Square	negative	14	2.00	29%	43%	29%
5	Red Square	positive	14	2.86	0%	14%	86%
5	Terracotta, zh	negative	11	2.27	18%	36%	45%
5	Terracotta, zh	positive	11	2.82	0%	18%	82%
5	Terracotta, en	negative	14	1.86	43%	29%	29%
5	Terracotta, en	positive	14	3.00	0%	0%	100%
6	Chichen Itza	negative	14	2.14	29%	29%	43%
6	Chichen Itza	positive	14	2.79	0%	21%	79%
6	Red Square	negative	14	2.07	21%	50%	29%
6	Red Square	positive	14	2.86	0%	14%	86%
6	Terracotta, zh	negative	11	2.00	18%	64%	18%
6	Terracotta, zh	positive	11	2.73	0%	27%	73%
6	Terracotta, en	negative	14	1.79	50%	21%	29%
6	Terracotta, en	positive	14	2.79	0%	21%	79%
7	Chichen Itza	negative	14	2.57	0%	43%	57%
7	Chichen Itza	positive	14	2.71	0%	29%	71%
7	Red Square	negative	14	2.07	29%	36%	36%
7	Red Square	positive	14	2.86	0%	14%	86%
7	Terracotta, zh	negative	11	2.00	27%	45%	27%
7	Terracotta, zh	positive	11	2.82	0%	18%	82%
7	Terracotta, en	negative	14	1.71	43%	43%	14%
7	Terracotta, en	positive	14	2.79	0%	21%	79%

Please help us study interpretability of the results from automated topic extraction!.

Box 1. Interpretability rating instructions.

This Excel file contains 106 word groups. The word groups represent topics in online reviews from Trip Advisor for three UN World Heritage sites. They were identified by applying a text mining probability-based computer algorithm (Latent Dirichlet Allocation) to visitors' reviews and need to be interpreted. We ask you to give an interpretability rating for each topic: 1 (poor), 2 (medium), or 3 (good).

- Interpretability of a topic is "good" if it contains words that can be grouped together as a single coherent concept. It is allowed for up to 3 words at the second half of the list not to be in the group. You can clearly name this topic.
- A topic is "medium" if it contains multiple unrelated concepts or if there are multiple words unrelated to a single coherent concept. You have trouble giving a single name to this topic.
- A topic is "poor" if it contains no clear, sensical connections between more than a few pairs of words. You cannot give a clear name to this topic.

Example

Interpretability 1: poor, 2: medium, 3: good	Topic words identified by computer algorithm	Interpretation (for your convenience only; do not write it)
3	terracotta warrior people horse time good feel spectacular history visit worth museum	All words connect into one topic describing a spectacular historical site well worth visiting. Possible topic name: "Description of museum, worth visiting"
2	tour guide attraction ticket station number holiday door student price summer train	Apparently there are several topics here: tours, ticket price (including for students), and how to get there by train. Possible topic name: "How to get there, get tour guide and a ticket"
1	soldier help full course audience move simple process solemn express style vibrant	Some word pairs make sense ("simple process", "full course [dinner]", "vibrant style"), but together the words are not clearly connected into a coherent concept. Hard to come up with the topic name.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tourman.2020.104241.

References

Au, N., Buhalis, D., & Law, R. (2009). Complaints on the online environment—the case of Hong Kong hotels. In W. Höpken, U. Gretzel, & R. Law (Eds.), *Information and communication technologies in tourism 2009* (pp. 73–85). Vienna: Springer-Verlag.

Au, N., Buhalis, D., & Law, R. (2014). Online complaining behavior in mainland China hotels: The perception of Chinese and non-Chinese customers. *International Journal of Hospitality & Tourism Administration, 15*(3), 248–274.

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, 243–268.

Beerli, A., & Martin, J. D. (2004). Factors influencing destination image. *Annals of Tourism Research, 31*(3), 657–681.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.

Borko, H., & Bernick, M. (1963). Automatic document classification. *Journal of the ACM, 10*(2), 151–162.

Breathnach, T. (2006). Looking for the real me: Locating the self in heritage tourism. *Journal of Heritage Tourism, 1*(2), 100–120.

Casaló, L. V., Flavián, C., Guinalíu, M., & Ekinci, Y. (2015). Avoiding the dark side of positive online consumer reviews: Enhancing reviews' usefulness for high risk-averse travelers. *Journal of Business Research, 68*(9), 1829–1835.

Cenni, I., & Goethals, P. (2017). Negative hotel reviews on TripAdvisor: A cross-linguistic analysis. *Discourse, Context & Media, 16*, 22–30.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Paper presented at the Advances in neural information processing systems*.

Chang, Y.-C., Ku, C.-H., & Chen, C.-H. (2019). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. *International Journal of Information Management, 48*, 263–279.

Cho, V. (1998). World wide web resources. *Annals of Tourism Research, 25*(2), 518–521.

Diamond, J. (1999). *Guns, Germs, and Steel: The Fates of human societies*. W. W. Norton.

Dinçer, M. Z., & Alrawadieh, Z. (2017). Negative word of mouse in the hotel industry: A content analysis of online reviews on luxury hotels in Jordan. *Journal of Hospitality Marketing & Management, 26*(8), 785–804.

Feinstein, E. (2018). *OTA's vs. direct hotel bookings: Which is the leading trend for 2018?*. Retrieved from https://www.traveldailynews.com/post/otas-vs-direct-hotel-bookings-which-is-the-leading-trend-for-2018.

Fong, L. H. N., Lei, S. S. I., & Law, R. (2017). Asymmetry of hotel ratings on TripAdvisor: Evidence from single-versus dual-valence reviews. *Journal of Hospitality Marketing & Management, 26*(1), 67–82.

Fuchs, M., Höpken, W., & Lexhagen, M. (2014). Big data analytics for knowledge generation in tourism destinations – a case from Sweden. *Journal of Destination Marketing & Management, 3*(4), 198–209.

Gülker, G., Hinsch, K. D., & Kraft, A. (2001). Deformation monitoring on ancient terracotta warriors by microscopic TV-holography. *Optics and Lasers in Engineering, 36*(5), 501–513.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management, 59*, 467–483.

Harrison-Walker, L. J. (2001). E-complaining: A content analysis of an internet complaint forum. *Journal of Services Marketing*.

Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. *Journal of Consumer Research, 17*(4), 454–462.

Heydt, M. (2018). *Python Web Scraping Cookbook: Over 90 proven recipes to get you scraping with Python, microservices, Docker, and AWS*. Packt Publishing.

Hosany, S., Ekinci, Y., & Uysal, M. (2006). Destination image and destination personality: An application of branding theories to tourism places. *Journal of Business Research, 59*(5), 638–642.

Huang, Y., Basu, C., & Hsu, M. K. (2010). Exploring motivations of travel knowledge sharing on social network sites: An empirical investigation of US college students. *Journal of Hospitality Marketing & Management, 19*(7), 717–734.

Hughes, M., & Carlsen, J. (2010). The business of cultural heritage tourism: Critical success factors. *Journal of Heritage Tourism, 5*(1), 17–32.

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management, 72*, 417–426.

Jacobsen, J. K. S., & Munar, A. M. (2012). Tourist information search and destination choice in a digital age. *Tourism Management Perspectives, 1*, 39–47.

Jiang, J., Gretzel, U., & Law, R. (2010). Do negative experiences always lead to dissatisfaction?–testing attribution theory in the context of online travel reviews. *Enter* (pp. 297–308).

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research, 57*(8), 1012–1025. https://doi.org/10.1177/0047287517729757

Kladou, S., & Mavragani, E. (2015). Assessing destination image: An online marketing approach and the case of TripAdvisor. *Journal of Destination Marketing & Management, 4*(3), 187–193.

Leung, D., Law, R., Van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing, 30*(1–2), 3–22.

Liao, H. (2007). Do it right this time: The role of employee service recovery performance in customer-perceived justice and customer loyalty after service failures. *Journal of Applied Psychology, 92*(2), 475.

Li, X., Pan, B., Zhang, L., & Smith, W. W. (2009). The effect of online information search on image development: Insights from a mixed-methods study. *Journal of Travel Research, 48*(1), 45–57.

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 Management, 59*, 554–563.

Li, Y., Ye, Q., Zhang, Z., & Wang, T. (2011). Snippet-based unsupervised approach for sentiment classification of Chinese online reviews. *International Journal of Information Technology and Decision Making, 10*(6), 1097–1110.

Locander, W. B., & Hermann, P. W. (1979). *The effect of self-confidence and anxiety on information seeking in consumer risk reduction*. Los Angeles, CA: SAGE Publications Sage CA.

Lu, H. (2019). *Qin Terracotta Army attracted more than 100 million visitors in past 40 years*. China Daily. Retrieved from http://www.chinadaily.com.cn/a/201909/25/WS5d8b0bc7a310cf3e3556d6b4.html.

Lu, W., & Stepchenkova, S. (2012). Ecotourism experiences reported online: Classification of satisfaction attributes. *Tourism Management, 33*(3), 702–712.

Lu, W., & Stepchenkova, S. (2015). User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. *Journal of Hospitality Marketing & Management, 24*(2), 119–154.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., et al. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures, 12*(2–3), 93–118.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge university press.

Marchant, R. (2015). *Local consumer review Survey 2015*. *Bright ideas,* October 19, 2015. Retrieved from https://www.brightlocal.com/research/local-consumer-review-survey-2015/.

Marine-Roig, E., & Clavé, S. A. (2015). Tourism analytics with massive user-generated content: A case study of barcelona. *Journal of Destination Marketing & Management, 4*(3), 162–172.

Maser, B., & Weiermair, K. (1998). Travel decision-making: From the vantage point of perceived risk and information preferences. *Journal of Travel & Tourism Marketing, 7*(4), 107–121.

Mauri, A. G., & Minazzi, R. (2013). Web reviews influence on expectations and purchasing intentions of hotel potential customers. *International Journal of Hospitality Management, 34*, 99–107.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Paper presented at the Proceedings of the conference on empirical methods in natural language processing*.

Moreira, P., & Iao, C. (2014). A longitudinal study on the factors of destination image, destination attraction and destination loyalty. *Journal of Social Sciences, 3*(3), 90.

Murphy, R. (2019). *2019 local consumer review Survey. Bright Ideas, December 11*. Retrieved from https://www.brightlocal.com/research/local-consumer-review-survey.

Newman, D. (2015). *Customer experience is the future of marketing*. Retrieved from https://www.forbes.com/sites/danielnewman/2015/10/13/customer-experience-is-the-future-of-marketing/#2f48c2f4193d.

Newman, D., Lau, J., Grieser, K., & Baldwin, T. (2010). *Automatic evaluation of topic coherence. InHuman language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics, HLT'10*. Stroudsburg, PA, USA: Association for Computational Linguistics.

Nielsen Global Online Consumer Survey. (2009). *Global advertising consumers trust real friends and virtual strangers the most*. Retrieved from https://www.nielsen.com/us/en/insights/article/2009/global-advertising-consumers-trust-real-friends-and-virtual-strangers-the-most/.

Newman, D., Noh, Y., Talley, E., Karimi, S., & Baldwin, T. (2010). Evaluating topic models for digital libraries. *Proceedings of the 10th annual joint conference on Digital libraries*.

Nonnecke, B., Andrews, D., & Preece, J. (2006). Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research, 6*(1), 7–20.

Olga, L., & Raj, R. (2013). Evolution of social media and consumer behavior changes in tourism destination promotion. *Internation Journal of Sales, Retailing & Marketing*, 31–39.

Pan, B., MacLaurin, T., & Crotts, J. C. (2007). Travel blogs and the implications for destination marketing. *Journal of Travel Research, 46*(1), 35–45.

Pantelidis, I. S. (2010). Electronic meal experience: A content analysis of online restaurant comments. *Cornell Hospitality Quarterly, 51*(4), 483–491.

Putri, I., & Kusumaningrum, R. (2017). Latent Dirichlet allocation (LDA) for sentiment analysis toward tourism review in Indonesia. In *Paper presented at the journal of physics: Conference series*.

Raschka, S. (2015). *Python machine learning*. Packt Publishing Ltd.

Rossetti, M., Stella, F., & Zanker, M. (2016). Analyzing user reviews in tourism with topic models. *Information Technology & Tourism, 16*(1), 5–21.

Schmunk, S., Höpken, W., Fuchs, M., & Lexhagen, M. (2013). *Sentiment analysis: Extracting decision-relevant knowledge from UGC Information and communication Technologies in tourism 2014* (pp. 253–265). Springer.

Shugan, S. M. (2007). The Anna Karenina bias: Which variables to observe? *INFORMS*.

Smith, A., & Anderson, M. (2016). *Online shopping and e-commerce* (Vol. 19). Pew Research Center.

Song, S., Kawamura, H., Uchida, J., & Saito, H. (2019). Determining tourist satisfaction from travel reviews. *Information Technology & Tourism, 21*(3), 337–367.

Sparks, B. A., & Browning, V. (2010). Complaining in cyberspace: The motives and forms of hotel guests' complaints online. *Journal of Hospitality Marketing & Management, 19*(7), 797–818. https://doi.org/10.1080/19368623.2010.508010%/Taylor&FrancisGroup

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management, 32*(6), 1310–1323. https://doi.org/10.1016/j.tourman.2010.12.011

Stankov, U., Lazic, L., & Dragicevic, V. (2010). The extent of use of basic Facebook user-generated content by the national tourism organizations in Europe. *European Journal of Tourism Research, 3*(2), 105–113.

Stepchenkova, S., Mills, J. E., & Jiang, H. (2007). Virtual travel communities: Self-reported experiences and satisfaction. In M. Sigala, L. Mich, & J. Murphy (Eds.), *Information and communication technologies in tourism 2007*. Springer. %\ 2020-01-14 09:58:00.

Stepchenkova, S., & Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism Management, 36*, 590–601. https://doi.org/10.1016/j.tourman.2012.08.006

Tang, J., Meng, Z., Nguyen, X., Mei, Q., & Zhang, M. (2014). Understanding the limiting factors of topic modeling via posterior contraction analysis. In *Paper presented at the.* International Conference on Machine Learning.

Tasci, A. D., Croes, R., & Villanueva, J. B. (2014). *Rise and fall of community-based tourism–facilitators, inhibitors and outcomes. Worldwide Hospitality and Tourism Themes*.

Tripadvisor. (2020). *About Tripadvisor*. Retrieved from https://tripadvisor.mediaroom.com/us-about-us.

UNESCO. (2020). *Mausoleum of the first Qin emperor*. Retrieved from http://whc.unesco.org/en/list/441.

UNWTO. (2019). *International tourism highlights*. Retrieved from https://www.e-unwto.org/doi/pdf/10.18111/9789284421152.

Valijärvi, R., & Tarsoly, E. (2019). "Language students as critical users of google translate": Pitfalls and possibilities'. *Practitioner Research in Higher Education Journal, 12*(1), 61–74.

Vásquez, C. (2011). Complaints online: The case of TripAdvisor. *Journal of Pragmatics, 43* (6), 1707–1717.

Windsor, L. C., Cupit, J. G., & Windsor, A. J. (2019). Automated content analysis across six languages. *PLOS One, 14*(11).

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 Management, 58*, 51–65.

Xiang, Z., Schwartz, Z., Gerdes, & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management, 44*, 120–130.

Yoo, K. H., & Gretzel, U. (2008). What motivates consumers to write online travel reviews? *Information Technology & Tourism, 10*(4), 283–295.

Yuksel, A., Kilinc, U., & Yuksel, F. (2006). Cross-national analysis of hotel customers' attitudes toward complaining and their complaining behaviours. *Tourism Management, 27*(1), 11–24.

Zheng, T., Youn, H., & Kincaid, C. S. (2009). An analysis of customers' E-complaints for luxury resort properties. *Journal of Hospitality Marketing & Management, 18*(7), 718–729. https://doi.org/10.1080/19368620903170240%/Taylor&FrancisGroup

Andrei P. Kirilenko (andrei.kirilenko@ufl.edu) is Associate Professor in the Department of Tourism, Hospitality and Event Management at the University of Florida. He received his Ph.D. in Computer Science and held positions at the Center for Ecology & Forest Productivity, Russia, European Forest Institute, Finland, U.S. Environmental Protection Agency laboratory, OR, Purdue University and University of North Dakota. His research interests include big data analysis, data mining, tourism analytics, climate change impacts, and sustainability issues.

Svetlana Stepchenkova (svetlana.step@ufl.edu) is Associate Professor at the Department of Tourism, Hospitality and Event Management at the University of Florida. Her research interests are in the area of marketing communications, branding, and positive image building. She studies tourism behavior and the effectiveness of destination promotion in situations of strained bilateral relations between nations. She is also interested in usability of user-generated content for managerial decision making in destination management.

Xiangyi Dai (realnae@126.com) is Associate Professor at the College of Resource Environment and Tourism at Capital Normal University. He received his PhD in Human Geography at the College of Urban and Environmental Sciences at Peking University. His research interests include cultural heritage conservation and utilization, cultural tourism, regional tourism development, and community participation. He is also a visiting scholar at the Department of Tourism, Hospitality and Event Management at the University of Florida during the year of 2020.