Journal of Destination Marketing & Management 20 (2021) 100598

Contents lists available at ScienceDirect

Journal of Destination Marketing & Management

journal homepage: www.elsevier.com/locate/jdmm

Exploring destination loyalty: Application of social media analytics in a nature-based tourism setting

Farshid Mirzaalian *, Elizabeth Halpenny

University of Alberta, Faculty of Kinesiology, Sport, and Recreation, Edmonton, Alberta, T6G 2H9, Canada

A R T I C L E I N F O

Keywords:
Nature-based tourism
Tourism experience
Destination loyalty
Use-generated content
Social media analytics
Sentiment analysis
Topic modeling
Text clustering

A B S T R A C T

User-generated content across social media platforms is playing an increasingly important role in tourism. Understanding tourists' experiences and opinions about tourism destinations has led to numerous opportunities to provide tourism providers with greater insights. Identifying sentiments, detecting topics of interest, and exploring loyalty behaviors from user-generated content can provide valuable direction for managerial decisions. This paper presents a novel and inclusive approach that uses different analytical techniques such as sentiment analysis and topic modeling to extract sentiments and topics of interest from tourists' conversational data on TripAdvisor from 2002 to 2019. Sentiment analysis revealed that some touristic locations in Jasper National Park are outperforming others in terms of sentiment scores, despite the fact that tourists less frequently reflect on their experiences at these places on social media. Such higher rankings suggest that average sentiment score can be a more informative measure than simple TripAdvisor rankings. This paper also explores destination loyalty statements using a keyword clustering approach. Previous destination loyalty literature was used to develop a keyword list that was applied to search for expressions of loyalty in online reviews. The robustness of loyalty clusters and optimal number of clusters was also assessed prior to final analysis. Four leading loyalty-focused categories of destination offerings were observed: glaciers, waterfalls, lakes and islands, and hiking and trails. Prioritization of visitor experience enhancements relating to these loyalty-inducing destination components are discussed.

1. Introduction

Social media (SM) has experienced tremendous growth in recent years, especially with the emergence of diverse SM platforms such as social network sites, discussion forums, wikis, picture and video sharing platforms, and ratings and reviews communities. In the tourism context, SM has also significantly revolutionized the way tourists seek information, plan their trips and, more importantly, share travel experiences with others. These different SM applications and platforms produce a remarkable amount of measurable data for destination marketers whose goal is to effectively render these data for decisions relating to promotion and offerings development (Buhalis & Law, 2008; Hays, Page, & Buhalis, 2013; Xiang & Gretzel, 2010). These different forms of user-generated content have not only enabled tourism actors to monitor and analyze tourists' behaviors and develop different marketing performance indicators but have also helped them communicate with consumers and plan long-term strategies (e.g. destination loyalty). Social media analytics (SMA) opens the door for destination marketing/management organizations (DMO) to develop new knowledge through reshaping their understanding of the field and making better business decisions with the use of decision support systems (Xiang, Schwartz, Gerdes, & Uysal, 2015).

Tracking the behavioral dynamics of tourists has become a major challenge for tourism destinations. DMOs and other tourism service providers in destinations such as tour operators are very interested in a number of factors. These include knowing the details of touristic locations that tourists visit, what factors attract the tourists to these locations, the tourists' subjective evaluations of the locations, their personal reflections and, most importantly, their loyalty behaviors such as future travel behavioral intentions and whether they will recommend the destination to others. Most current research advances are not capable of addressing these issues with a decision-centric, integrated and comprehensive approach. In fact, most of the existing approaches for studying SM data are focused on tackling nonexploratory questions that are already predefined and rarely assist in generating understanding of tourists' interests, emotions, experiences, and loyalty behaviors (Miah,

* Corresponding author.
E-mail addresses: farshid.mirzaalian@ualberta.ca (F. Mirzaalian), elizabeth.halpenny@ualberta.ca (E. Halpenny).

https://doi.org/10.1016/j.jdmm.2021.100598
Received 14 February 2020; Received in revised form 10 March 2021; Accepted 13 March 2021
Available online 23 March 2021

Vu, Gammack, & McGrath, 2017). This study seeks to develop and evaluate a new analytics method based on textual content of tourists' online reviews about Jasper National Park (JNP) as a nature-based tourism destination. This study incorporates emerging computational methods to provide a management-driven framework in which the details of the proposed design artefact are specified as a nature-based tourism destination management strategic planning and operational decision support tool. In order to come up with a more effective solution, this study brings together four computational techniques (text processing, sentiment analysis, latent dirichlet allocation topic modeling, and text clustering) to more comprehensively tackle the DMO's decision-support needs. Combined, these methods have the capacity to provide insight into tourists' loyalty behavior to support DMOs with tourism development, management and planning.

2. Literature review

2.1. Destination loyalty

Loyalty has become a critical part of tourism research in recent decades. Tourism providers realize the importance of loyal visitors, knowing that their competitors offer similar attractions, services and experiences. Destination managers try to maintain an acceptable level of service and maximize visitor satisfaction within given constraints. To convert visitors to loyal patrons, destinations first need to know what visitors' expectations are, so that they can meet and potentially exceed those expectations by providing appealing services before, during and after their visit. Understanding how visitors form their destination loyalty and what factors influence their loyalty formation is important for the success of tourism destinations.

There are three main approaches for defining and measuring tourist loyalty: measuring attitude, measuring behavior, or measuring a combination of two. The behavioral perspective focuses on a tourist's actual consumption behavior such as repeat visit duration, frequency and intensity (Oppermann, 2000). In contrast to the behavioral approach that produces only a static outcome of a dynamic process, the attitudinal perspective goes beyond and considers loyalty in terms of tourists' strength of affection toward a destination or attraction (Pritchard & Howard, 1997). Finally, a composite conceptualization of loyalty integrates both behavioral and attitudinal dimensions, by not only looking at the tourist's consumption behavior such as repeated visits, but by considering future actions such as willingness to recommend to third parties (Oppermann, 2000), the strength of preference (Lee, Yoon, & Lee, 2007b), and the feeling of attachment towards the place (Yuksel, Yuksel, & Bilim, 2010). Chen and Gursoy (2001) argued that a composite measure of loyalty (combination of both attitudinal and behavioral measures) provides the most accurate representation of destination loyalty. Identifying determinants of loyalty has been an important research topic among tourism researchers. While some loyalty-related researchers have focused on factors such as activity (Backman & Crompton, 1991), service quality (Baker & Crompton, 2000), and tourism providers (Morais, Dorsch, & Backman, 2004), other researchers have pointed out the importance of commitment to a specific place, or what is referred to as destination loyalty (Kyle, Graefe, Manning, & Bacon, 2004; Oppermann, 2000).

2.2. Antecedents of tourist loyalty

2.2.1. Service quality and satisfaction

There is a general agreement about the positive relationship between service quality and satisfaction, and that quality service and satisfaction can lead to loyalty (Baker & Crompton, 2000; Mason & Nassivera, 2013). There is also evidence of a mediatory effect of tourist satisfaction in the relationship between service quality and behavioral intentions (Chen & Chen, 2010). Moreover, satisfaction directly affects destination choice (Tian-Cole & Cromption, 2003), revisit intentions (Um, Chon, & Ro, 2006), and recommendations to others (Lee, Yoon, & Lee, 2007b). Tourist satisfaction is one of the most commonly used determinants of loyalty and plays an important role in the success of a tourism destination.

2.2.2. Destination image

Destination image can be generally defined as a person's collection of beliefs, impressions, benefits, and attributes of a destination based on information he or she has gradually processed from various sources (Zhang, Fu, Cai, & Lu, 2014). Destination image plays an important role in the tourists' decision-making processes, from pre-visit planning to post-visit consequent behaviors (i.e. did they complain to friends or praise the place). Previous studies have found positive relationships between image and satisfaction and image and quality (Chen & Tsai, 2007; Chi & Qu, 2008). These relationships can indirectly influence loyalty. At the same time, a strong relationship has also been found between destination image and behavioral intentions of tourists, such as the intention to revisit the same destination in the future (Kim, Hallab, & Kim, 2012).

2.2.3. Travel motivations

Travel motivation is one of the first steps in the travel decision-making process of tourists and has been widely examined in the tourism context. Drawing from social psychological theories (Iso-Ahola, 1982) and expectancy theories (Witt & Wright, 1992), the push-pull theory (Gavcar & Gursoy, 2002) has been one of the frequently utilized approaches to study the reasons that people travel. This theory suggests that the main motives for tourists to travel to a destination are push factors or what people expect from the travel experience (i.e. personal internal desires such as the opportunity to escape from the daily routine or the opportunity to have fun), and pull factors or what the destination offers to create the experience (i.e. destination-specific attributes and attractions such as entertainment opportunities and good quality restaurants) (Gursoy, S.Chen, & G.Chi, 2014; Prebensen, Woo, Chen, & Uysal, 2013). Previous research suggests that push and pull significantly impact tourists' motivations with regards to their level of satisfaction with a destination, which in turn indirectly influences their loyalty (Prebensen et al., 2013; Yoon & Uysal, 2005).

2.2.4. Previous experience and involvement

Previous experience is the number of previous visits to a tourism destination and the length of these experiences. This measure, along with place attachment or the level of involvement with a destination, can affect destination image formation, revisit intention, and positive recommendations to others (Beerli & Martin, 2004; Milman & Pizam, 1995; Sönmez & Graefe, 1998). Previous experiences inspire tourists to encourage cognitive, affective and conative ties with a destination compared to a traveler with fewer or no previous trips (Gursoy et al., 2014; Halpenny, Kulczycki, & Moghimehfar, 2016; Yuksel et al., 2010).

The level of involvement with a destination (also known as 'tourist personal relevance') plays an important role as a direct antecedent of tourist loyalty. Tourists' level of involvement depends on the degree to which the destination characteristics match the tourists' expectations, goals, values, and their prior knowledge of the destination (Gursoy et al., 2014). Personal involvement has been also defined as the degree to which tourists devote themselves to an activity or experience (Zaichkowsky, 1985). In tourism research, evidence for the relationship between personal involvement and tourist loyalty is mixed. This evidence comes from previous studies showing that different aspects of involvement can have different forms of influence on behavioral constructs such as satisfaction and loyalty (Prayag & Ryan, 2012; Sparks, 2007).

2.2.5. Place attachment

Place attachment refers to the nature and nuances of a tourist's emotional relationship with places and tourism destinations. It has two main components: place dependence and place identity (Williams &

Roggenbuck, 1989). Place dependence is referred to as the level of attachment to a place because of its functional reasons and the use of its resources. Place identity is defined as an individual's value judgment in relation to the place, influenced by emotional developments over time such as beliefs, preferences, feelings, values, goals, etc. (Williams & Roggenbuck, 1989).

Place attachment as an antecedent variable has been widely studied in hospitality and tourism, as a predictor of tourist satisfaction and tourist loyalty, spending preferences, pro-environmental behavior, and leisure participation patterns (Alexandris, Kouthouris, & Meligdis, 2006; George & George, 2004; Halpenny, 2006; Lee, Graefe, & Burns, 2007a). Previous research shows that place attachment and destination loyalty are significantly and positively related (Lee, Graefe, & Burns, 2007a; Yuksel et al., 2010).

2.2.6. *Perceived value*

Perceived value is rooted in equity theory. Under the equity concept, consumers compare their monetary payments or nonmonetary investments such as time and energy with the output they receive from providers, and evaluate what is fair, right, or deserved for the perceived value (Bolton & Lemon, 1999). Perceived value and its effects on consumers' perception about quality of services, satisfaction and destination loyalty has been extensively studied in the hospitality and tourism context, suggesting that perceived value is a key determinant of satisfaction, perception of quality and loyalty behavior (Chen & Chen, 2010; Gallarza & Saura, 2006; Sun, Chi, & Xu, 2013; Velázquez, Saura, & Molina, 2011).

The related destination loyalty literature cited above was applied to develop a keyword list that can be used to detect destination loyalty expressions and statements that visitors posted on TripAdvisor about Jasper National Park. These keywords include but are not limited to: 'revisit,' 'visit again,' 'come back,' 'recommend,' and 'worth.' (See section 3.4 for the full list) Next, an introduction to social media analytics (SMA) is provided, followed by an overview of current methods for evaluating destination loyalty expressed through social media posting.

2.3. *Social media analytics*

SMA refers to a variety of interdisciplinary techniques and informatics tools such as Web crawling, computational linguistics, machine learning, and statistical methods to "collect, monitor, analyze, summarize, and visualize SM data, usually driven by specific requirements from a target application'' (Zeng, Chen, Lusch, & Li, 2010, p. 14). In a systematic literature review on the applications of SMA in hospitality and tourism, Mirzaalian and Halpenny (2019) reported that the majority of tourism studies have applied SMA to explore destination image (Fuchs, Höpken, & Lexhagen, 2014; Költringer & Dickinger, 2015; Li, Lin, Tsai, & Wang, 2015), destination satisfaction (Capriello, Mason, Davis, & Crotts, 2013; Del Vecchio, Mele, Ndou, & Secundo, 2018), and travel patterns/tourist flow (Chua, Servillo, Marcheggiani, & Moere, 2016; Vu, Li, Law, & Ye, 2015; Zhou, Xu, & Kimmons, 2015). Tourism studies have also used SMA to predict destination visits (Miah et al., 2017; Pantano, Priporas, & Stylos, 2017) and measure the performance and test the accuracy of analytical methods (Kirilenko, Stepchenkova, Kim, & Li, 2018; Ye, Zhang, & Law, 2009). The variety of SM analytical methods that have been used in tourism studies include but are not limited to text analytics, clustering and topic modeling, sentiment analysis, trend analysis, predictive analytics, and spatial analysis (Mirzaalian & Halpenny, 2019; Stieglitz & Dang-Xuan, 2013).

Positive and negative online reviews are full of insights that help tourism providers to understand brand value in the mind of consumers, and whether they have been able to deliver their brand promise. Sentiment analysis of negative reviews, for instance, highlights where a destination has failed to deliver services that were claimed in its mission, while on the other end of the spectrum, analyzing the most enthusiastic reviewers from loyal visitors can give DMOs ideas as to how to reach more visitors while reinforcing revisit intentions among loyal tourists. Analyzing a visitor's sentiment expressed in online reviews is also important for DMOs to have an informed understanding of the experience and subjective opinions of visitors toward the destination, detailed insight that would not be gained by relying only on comments and the overall experience rating.

Topic identification (also known as feature extraction) is another useful method which focuses on extracting features of a specific product or service and distinguishes the topic of online reviews by assigning a predefined topic (supervised machine learning techniques) or identifying unknown topics (not predefined) mentioned within a review statement (unsupervised method). The latter turns out to be a promising approach in the tourism context, specifically for tourism destinations, to gain new insights into 'not previously recognized' relevant quality dimensions of tourism services, as well as strengths and weaknesses of concrete tourism services along those quality dimensions (Menner, Höpken, Fuchs, & Lexhagen, 2016).

SM platforms can be categorized into social networking sites (e.g. Facebook, Twitter), discussion forums (e.g. TripAdvisor Travel Forum), media and content communities (e.g. Flickr, YouTube), and consumer review sites (e.g. TripAdvisor, Yelp) (Mirzaalian & Halpenny, 2019). Social networking sites refers to web-based applications and services where public or semi-public users can connect with each other and share similar personal interests, lifestyles, or activities based on the nature of the site (Boyd & Ellison, 2007), while discussion forums are mainly organized around people with common interests where they can share their knowledge and experience in different areas. Media and content communities refer to web and mobile applications which enable their users to share content such as photos and videos. Finally, consumer review sites refer to platforms on which consumers can post content about products and services. The majority of SM analytical research in the hospitality and tourism context has focused on consumer review sites (specifically on TripAdvisor and Daodao.com), social networking sites (explicitly Twitter and Sina Weibo), and media/content communities (Flickr in particular) (Mirzaalian & Halpenny, 2019). Consumer review sites (also referred to as 'consumer-generated media') in the hospitality and tourism context can be categorized into community-based websites and transaction-based online travel agencies (Gligorijevic, 2016). In the former case, online platforms such as TripAdvisor combine a variety of user data, information tools, and travel forums to represent different aspects of destinations (or hotels and restaurants), while the focus in transaction-based platforms such as Expedia and Bookings is more on financial aspects of tourism (Xiang, Du, Ma, & Fan, 2017). Differences between these two data sources must be considered for the accuracy, representativeness, and quality of data in SM research in general, and tourism-related online reviews in particular. For instance, gathered data about a specific destination from social networking sites such as Facebook and Twitter are unstructured in nature, which makes the interpretation challenging, while exploring structured data collected from other online travel websites like TripAdvisor is more practicable.

TripAdvisor is one of the largest travel sites, the world's largest travel community, with an average of 455 million unique visitors every month. It generated approximately 730 million user reviews and opinions covering more than eight million listings for restaurants, hotels, vacation rentals and attractions (Statista, 2019). TripAdvisor has a unique feature of 'Top Things to Do' for each specific tourism destination. This feature provides classified review-based information for the entire destination. Travelers can limit their search results based on different criteria and 'Types of Attractions' such as 'Nature and Parks,' 'Outdoor Activities,' 'Sights and Landmarks,' etc. This destination-based feature has made TripAdvisor an appealing avenue for hospitality and tourism studies, especially for outdoor tourism destinations such as national parks and natural attractions. For example, in a study of 5000 TripAdvisor reviews of 843 hotels, relationships between sentiment, rating, volume and variation of reviews and hotel performance were examined;

results revealed that overall and specific ratings, variation and volume of reviews, and the number of management responses were associated with hotel performance (Xie, So, & Wang, 2017). Another study of 373 TripAdvisor reviews of Costa Rica Ecolodges used exploratory content analysis and linear regression to find influential factors on ecotourists' satisfaction (Lu & Stepchenkova, 2012). Their quantitatively supported method classified satisfaction attributes into satisfiers, dissatisfiers, critical, and neutrals. Pearce and Wu (2018) also used an exploratory content analysis of 350 TripAdvisor reviews of entertainment performances at a China-based attraction. Their findings suggest that international tourists were generally positive toward the entertainment while sharing their experiences in TripAdvisor and were mainly attracted to the attraction's culturally distinctive style (Pearce & Wu, 2018). Another study examined 20,000 TripAdvisor reviews of 106 attractions in New Orleans. Using review readability, reviewer characteristics, and review rating, the authors examined which factors led people to judge a review as helpful. The results showed that review readability and reviewer characteristics are the most influential factors that affect the perceived value of reviews (Fang, Ye, Kucukusta, & Law, 2016).

2.4. Evaluating destination loyalty construct on social media

Social media has fundamentally revolutionized the way tourists communicate, collaborate, consume, and generate information related to destinations. SM also characterizes one of the most transformative impacts of information technology on tourism, both within and outside destination boundaries. Previous tourism studies tried to demonstrate different antecedents of loyalty including satisfaction, service quality, perceived value, and communication through a variety of survey research methods. While very useful for identifying relationships, one potential constraint of survey research is that variables are defined by the researcher. User-generated content and electronic word of mouth (eWOM) posted by frequent travelers on travel websites and Internet forums such as TripAdvisor.com provide a rich source of self-reported, publicly accessible, unconstrained data, enabling researchers to enter the minds of tourists without any set parameters and explore their true thoughts on loyalty (Berezan, Raab, Tanford, & Kim, 2015). Obtaining market research data and understanding social interaction from online communities, what is referred to as netnography, is considered an efficient and naturalistic method of data collection. It has been argued that this method can outperform traditional data collection methods (e.g. focus groups, interviews), as it is spontaneously generated by consumers and thus reflects perceptions that are not easily obtained through other means. A benefit of this unsolicited content is that people may be more open and honest online than in face-to-face situations (Kozinets, 2002; Reid, 1996).

Social media has also provided a new marketing opportunity for hospitality and tourism providers to create interactive relationships with consumers. This shift from offline activities to online is an influential factor in building customers' loyalty (Senders, Govers, & Neuts, 2013). The eWOM posted by tourists in their different stages of travel (i.e. before, during, and after trip) has an influential effect on the reputation of tourism destinations. Therefore, providing timely feedback on user-generated content is becoming more and more important for suppliers and DMOs to build tourist trust, attract potential tourists, and encourage return visitation (McKay, Van Winkle, & Halpenny, 2019; Zeng & Gerritsen, 2014). Destinations can see the impact of their retention and loyalty efforts and identify opportunities for improvement by analyzing online reviews and user-generated content. Exploring themes of online reviews can help suppliers and DMOs recognize visitors' expectations and understand if they are met.

However, some argue that online reviews are inherently incomplete since they fail to reflect the opinions of users who have different propensities to post a review (Hargittai, 2020), or those with differing sentiments toward their experience (Chen, Zheng, & Ceran, 2016). Overlooking these silent users can result in a reporting bias (Chen et al., 2016; Hargittai, 2020; Morstatter & Liu, 2017). Moreover, hospitality and tourism online reviews tend to be more positive in nature in comparison with other service industries such as banking and finance (Gilbert & Veloutsou, 2006), mainly because expressing negative feelings is not an important motive behind writing reviews for tourists (Yoo & Gretzel, 2008). For the above-mentioned reasons, SM data has to be treated with caution and researchers should be aware of such potential biases when applying study results beyond particular online groups. Combining SM data with data collected from other traditional methods (e.g. interviews, focus groups, surveys) may be useful if the researcher seeks to generalize to groups other than the populations studied (Kozinets, 2002). These considerations helped to guide the research design employed in this current study. Explanations of the four analytics methods used in this paper and research procedures are provided in the following methodology section.

3. Methodology

All travelers' online reviews about top natural attractions and park areas in JNP were extracted from the third-party review website TripAdvisor, ranging from as early as December 2002 to October 2019. The reviews were collected in October 2019 (a total of 19807 reviews). Non-English reviews made up less than 15% of the corpus, however, only English reviews (17,224 reviews) were included for further analysis to avoid misinterpretation of comments written in other languages. Moreover, a loyalty keyword vocabulary was developed and employed in this study that contained English terms, therefore, only English reviews could be identified and separated from the rest of the corpus. JNP is the largest national park in the Canadian Rockies and part of UNESCO's Canadian Rocky Mountain Parks World Heritage Site (Parks Canada, 2019). Top natural attractions and park areas listed by TripAdvisor are as follows: Annette Lake, Athabasca Falls, Athabasca Glaciers, Columbia Icefield, Maligne Canyon, Maligne Lake, Mt. Edith Cavell, Mt. Edith Cavell Trail, Pyramid and Patricia Lakes, Spirit Island, Sulphur Skyline Trail, and Sunwapta Falls and Canyon. What follows is an explanation for the four analytic steps used to explore the extracted online reviews: text processing, sentiment analysis, latent dirichlet allocation topic modeling, and text clustering.

3.1. Text processing

For data extraction, the client-side software and data extraction tool Octoparse was used, which extracts web data through the application of advanced machine learning algorithms (Octoparse, 2019). Online reviews from TripAdvisor were extracted for this study. The extracted text-based online reviews were pre-processed and prepared for further analysis using four analytic methods. Some common pre-processing steps were splitting reviews into sentences through regular expressions based on punctuation (e.g. exclamation points, question marks), and splitting sentences into words (tokenization). Further pre-processing steps were stop-words removal (e.g. 'the,' 'a,' 'and'), stemming (e.g. removal of suffixes e.g. 'ing'), part-of-speech (POS) tagging (e.g. identification of words as nouns, verbs, adjectives, etc), and lowercase transformation. Full reviews or single sentences were finally transformed into a term-document-matrix, which describes the frequency of terms that occur in each respective posting. This transformation is based on term occurrences, term frequency, and inverse document frequency values.

3.2. Sentiment analysis

Computer-assisted sentiment analysis has unique advantages such as outperforming manual content coding analyses in terms of efficiency and reliability of the results (Capriello et al., 2013), and also significantly lowers cost, time, and labor compared to traditional methods like surveys or focus groups (Chiu, Chiu, Sung, & Hsieh, 2015). Sentiment analysis is premised on the idea that the content of a review is based either on opinions, personal feelings, beliefs, and judgment about entities or events (i.e. subjective), or is based on facts, evidence, and measurable observations (i.e. objective) (Feldman, 2013). In the case of tourism, online reviews and SM posts often reflect tourists' (dis)satisfaction, happiness, frustration, or disappointment toward tourism products and destinations (Schuckert, Liu, & Law, 2015). Sentiment analysis can be performed through either supervised technique or unsupervised (lexicon-based approach). Although showing a relatively higher performance than other methods (Chaovalit & Zhou, 2005; Kirilenko et al., 2018), supervised machine learning techniques have not been widely applied in tourism research (Mirzaalian & Halpenny, 2019). Therefore, there is a need for tourism studies to apply accuracy testing and report performance measurements of methods (Ye et al., 2009) to evaluate the robustness of sentiment analysis (See Appendix C for more information about sentiment analysis and differences between approaches). This study employed a supervised machine learning approach, where online reviews were first categorized into positive, neutral, and negative using the Valence Aware Dictionary and sEntiment Reasoner (VADER) sentiment (Gilbert & Hutto, 2014), followed by classification of the corpus into training set and test set to predict sentiments and evaluate accuracy of the prediction model (See Fig. 1 for supervised sentiment analysis procedure). 2-fold cross-validation was conducted to examine 15,972 positive and 920 negative reviews (i.e. Cross-validation is a statistical method and resampling procedure used to evaluate machine learning models on a limited data sample, where the parameter K refers to the number of groups that a given data sample is to be split into). The sentiment score was constructed by scoring the online reviews for positive and negative terms, and was calculated by adding a point for each positive word to the total score and deducting a point for each negative word (no points were given for neutral words) (Miner et al., 2012; Philander & Zhong, 2016). The Pos/Neg ratio score is computed as the ratio of overall positive words in each location to overall negative words, with any neutral word discarded. The average number of words in any of 12 attractions have been also reported.

3.3. LDA topic modeling

The second analytic approach was latent dirichlet allocation (LDA) topic modeling (Blei, Ng, & Jordan, 2003), which was used to effectively extract dimensions of the visitor experience from the corpus of text data extracted from TripAdvisor. Topic modeling is a good method for finding hidden semantic structures of online reviews and discovering the main topics and meaningful dimensions of visitors' experience-sharing regarding JNP. LDA is the most common method for topic modeling and is a generalization of probabilistic latent semantic indexing (PLSI) (Hofmann, 1999) (See Appendix A for further information about LDA topic modeling). LDA model was adopted instead of other text classification methods mainly because LDA modeling not only surpasses other methods in efficiently analyzing large-scale data at a highly granular level, but because it also helps to clarify the practical frequency of occurrence of each extracted dimension based on its intensity in the corpus (Guo, Barnes, & Jia, 2017). Revealed topics represent the important aspects related to tourists' experience and have a distribution across the online reviews depending on their frequency of occurrence. Over the last decade a number of improvements have been made to evaluate the semantic interpretability of topics and topic coherence within SM posts (Chang, Boyd-Graber, Wang, Gerrish, & Blei, 2009; Lau, Newman, & Baldwin, 2014; Newman, Lau, Grieser, & Baldwin, 2010) (See Appendix B for further background). The output from these LDA processes result in topics ranked by 'coherence level.' In this study, the Elbow Method was applied during the LDA modeling to examine the coherence value (i.e. the degree of semantic similarity between high scoring words in the topic), and to determine the appropriate number of topics for LDA model (Ketchen & Shook, 1996) (Fig. 3).

3.4. Text clustering

Text clustering is the application of cluster analysis to textual documents and is the process of finding groups of similar objects in the text, where the objects to be clustered can be documents, paragraphs, sentences or terms (Aggarwal & Zhai, 2012). Text clustering is a widely studied method used for data mining on textual contents, which through using different feature extraction techniques, sentences are converted into a term-document-matrix (Menner et al., 2016; Pang, Lee, & Vaithyanathan, 2002). One of the commonly used feature extraction techniques, based on term occurrences, is called Term Frequency (TF) or Term Frequency–Inverse Document Frequency (TF-IDF). A term document matrix with the TF-IDF weighted review words represents the basis of the k-means clustering algorithm, which was used with the cosine similarity (i.e. similarity between two non-zero vectors) as a distance measure as highly recommended by the text mining literature (Schuckert et al., 2015). Words with high TF-IDF values within a cluster then represent words often co-occurring in online reviews and represent latent topics (See Appendix D for further information about text clustering). This paper mainly adapts a clustering approach used in the Menner et al. (2016) study, where the authors utilized a term-document-matrix to identify relevant topics in tourism online reviews by performing a keyword clustering based on TF-IDF values. Therefore, it was assumed that topics are typically represented by special parts of speech, and that important words of an online review represent the major topics of that review. Therefore, frequent nouns have been extracted as topics, while frequent verbs have also been treated as topic words (Wartena & Brussee, 2008).

That being explained, after a detailed review of the hospitality and tourism loyalty literature by the author, a vocabulary of destination loyalty keywords was developed. This keyword vocabulary was subsequently used to identify and separate loyalty-expressed reviews from the rest of the corpus (i.e. Loyalty keyword vocabulary: 'revisit,' 'visit again,' 'return,' 'go back,' 'come back,' 'must,' 'recommend,' 'repeat,'

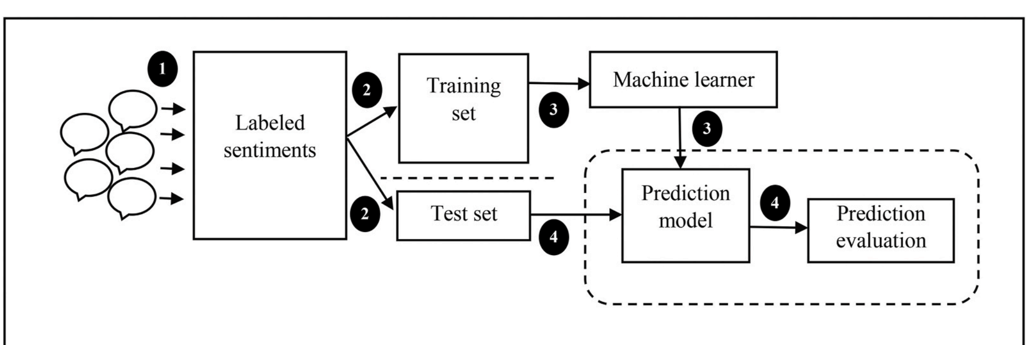

Fig. 1. Supervised sentiment analysis procedure.

'refer,' 'worth,' 'again,' 'loyal'). A close reading check was further applied to ensure the relatedness of the extracted reviews to destination loyalty conversations. The elbow method was applied in order to select the optimal number of clusters (i.e. 4 clusters), where based on the nature and characteristics of each location, destination loyalty expressions were categorized into 4 main predefined labeled clusters and were prepared for a more sophisticated supervised learning clustering (e.g. Athabasca Falls into waterfalls, Athabasca Glaciers into glaciers, Annette Lake into lakes and Islands, and Sulphur Skyline Trails into hiking and trails). A term-document-matrix with the TF-IDF weighted review words was used for this clustering approach, where the matrix characterizes the basis of the k-means clustering algorithm with the cosine similarity as distance measure as recommended by the text-document-clustering literature (Huang, 2008; Menner et al., 2016). Words with high TF-IDF values within a cluster then represent words often co-occurring in loyalty-expressed reviews and, thus, represent topics. A summary of what explained above and different steps toward the text clustering task is shown in Fig. 2.

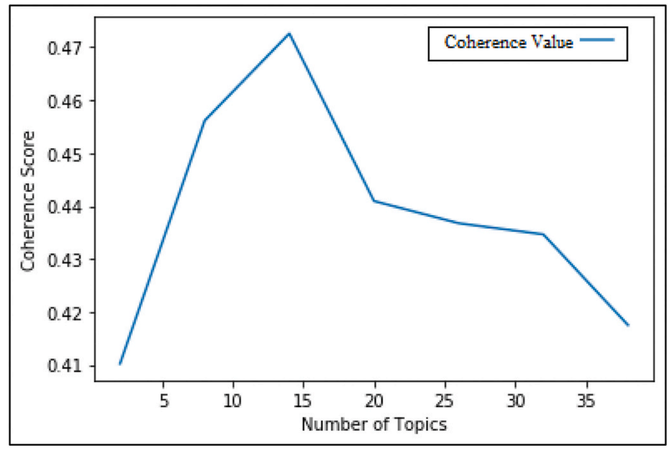

Fig. 3. Coherence values based on number of topics.

4. Results

In this section, results of the extraction of the dimensions of tourism experience, sentiment analysis and subjective evaluation of tourists' opinions, and dimensions of tourists' destination loyalty are summarized. The validity of these dimensions was then examined through reporting the performance measures of applied methods.

4.1. Sentiment analysis

Sentiment scores of the 12 touristic locations of JNP are provided in Table 1; special attention should be directed to the 'Average sentiment rank' and 'TripAdvisor relative rank' columns. The Pos/Neg ratio ranking, for instance, is calculated by dividing the overall Pos ratio scores to Neg ratio scores of total reviews in each location. For example, Annette Lake has a Pos/Neg ratio score of 17.65 (i.e. 0.226 overall positive ratio score to 0.013 overall negative ratio score of 87 reviews, neutral ratio discarded), and is ranked first based on this particular measure amongst other sights. In general, positive online reviews significantly outweighed negative reviews. This aligns with the observations (Yoo & Gretzel, 2008) that hospitality and tourism online reviews tend to be more positive in nature because expressing negative feelings is not an important motive behind writing reviews, especially in comparison with other service industries (e.g. banking and finance) which have a lower rate of positive reviews (Gilbert & Veloutsou, 2006).

On average, there were 1435 online reviews per location, with Athabasca Falls having the most reviews (4319), and Annette Lake the fewest (87). Some locations with lower review volumes, such as Sulphur Skyline Trail and Spirit Island, appeared in the top five average sentiment score rank, as along with other well-known locations with higher review volumes such as Maligne Lake. The average number of words in 12 attractions shows that despite significant differences between some attractions in terms of overall volume of reviews, all 12 documents are to some extent consistent in terms of average number of words used in online reviews. Also, although the average sentiment score and the Pos/Neg ratio score ranks were aligned with one another for most of the locations, some attractions had meaningfully different ranks such as Sulphur Skyline and Mt. Edith Cavell Trails, Spirit Island, and Athabasca Falls. Another remarkable finding upon comparing TripAdvisor relative rank with sentiment and ratio scores is that lakes and islands are relatively ranked lower on TripAdvisor in contrast to higher sentiment and ratio ranks uncovered in the results. These attractions are Annetee Lake, Pyramid and Patricia Lakes, Maligne Lake, and Spirit Island.

4.1.1. Performance evaluation of sentiment analysis

Accuracy, Precision, Recall, and F1-score were used for evaluating the results of sentiment analysis. Accuracy measures how accurate the method is in its prediction task through dividing the number of correct predictions by the total number of predictions. Precision is defined as the ratio of the number of cases correctly classified as one of the Pos, Neg, or Neu classes relative to the total number of cases predicted as that class. Respectively, the Recall of a class is defined as the relative number of

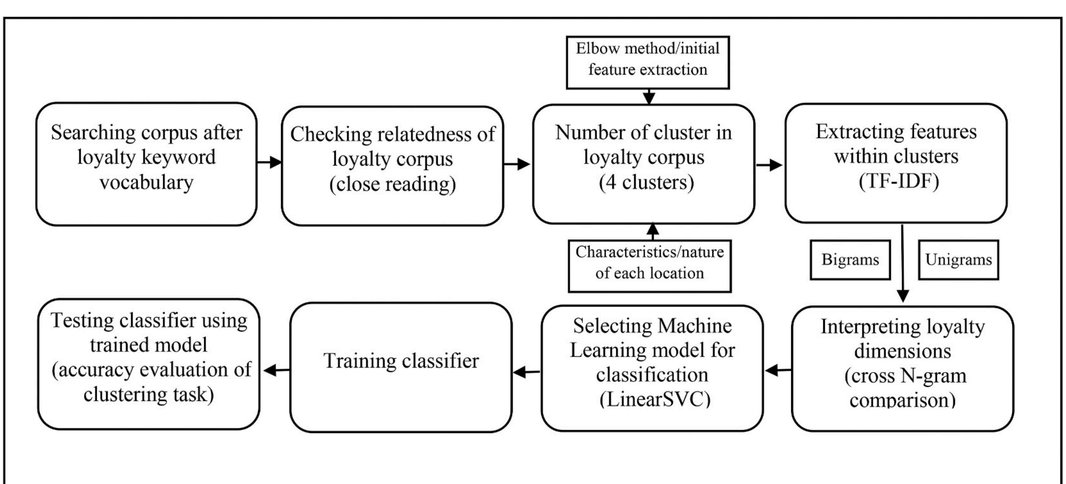

Fig. 2. Summary of strategy and different steps followed in the text clustering task.

Table 1
Attractions' sentiment scores.

	Review volume	≈Avg number of words	Avg sentiment score	Avg sentiment rank	Pos/Neg ratio score	Pos/Neg ratio rank	TripAdvisor relative rank
Sulphur Skyline Trail	134	118	0.82	1	6.29	9	2
Mt. Edith Cavell Trail	168	101	0.80	2	6.39	8	1
Annette Lake	87	56	0.79	3	17.65	1	7
Maligne Lake	1110	74	0.76	4	9.03	4	9
Spirit Island	240	70	0.75	5	12.02	2	6
Pyramid & Patricia Lakes	1425	50	0.74	6	11.25	3	10
Maligne Canyon	3740	55	0.71	7	8.47	5	3
Mt. Edith Cavell	472	87	0.71	8	6.49	7	5
Athabasca Falls	4319	41	0.70	9	6.84	6	4
Athabasca Glaciers	709	90	0.69	10	5.81	10	8
Sunwapta Falls & Canyon	595	52	0.65	11	4.39	12	12
Columbia Icefield	4225	71	0.64	12	5.51	11	11

cases correctly classified as one of the classes compared to the total number of instances. Finally, the F1-score is a weighted harmonic mean of both, the Precision and Recall. Results of the classification report for the sentiment rating shows an acceptable level of evaluation for the prediction model in each class, except for Neu most likely due to its smaller sample compared to other 2 classes (Pos and Neg), as well as satisfying weighted average for all 3 classes (93% for Accuracy, 87% for Precision, 93% for Recall, and 90% for F1-score).

4.2. Dimensions of tourism experience

LDA was applied to extract and label the dimensions of tourist experience across all collected online reviews from top touristic locations. LDA identified 14 topics and within each topic showed the top-20 words and their relative weight.

The labeling of dimensions was first conducted by one researcher and then confirmed by a second researcher. Labeling was based on the identification of a logical connection between the most frequent words for a topic. For example, in Figs. 4 and 5, a sample of 4 topics with word cloud and relative weights of their top 10 words has been shown. As an example, the topic labeled as 'Water/Cruise Tour' is based on the words 'experience,' weighted 61%, 'cruise' (40%), 'great' (35%), and 'trip' (26%), all of which appear at the top of the list (see Figs. 4 and 5). Another example is the topic labeled as 'Waterfalls Visit Experience'

Fig. 5. Word cloud of sample topics with their top 10 words.

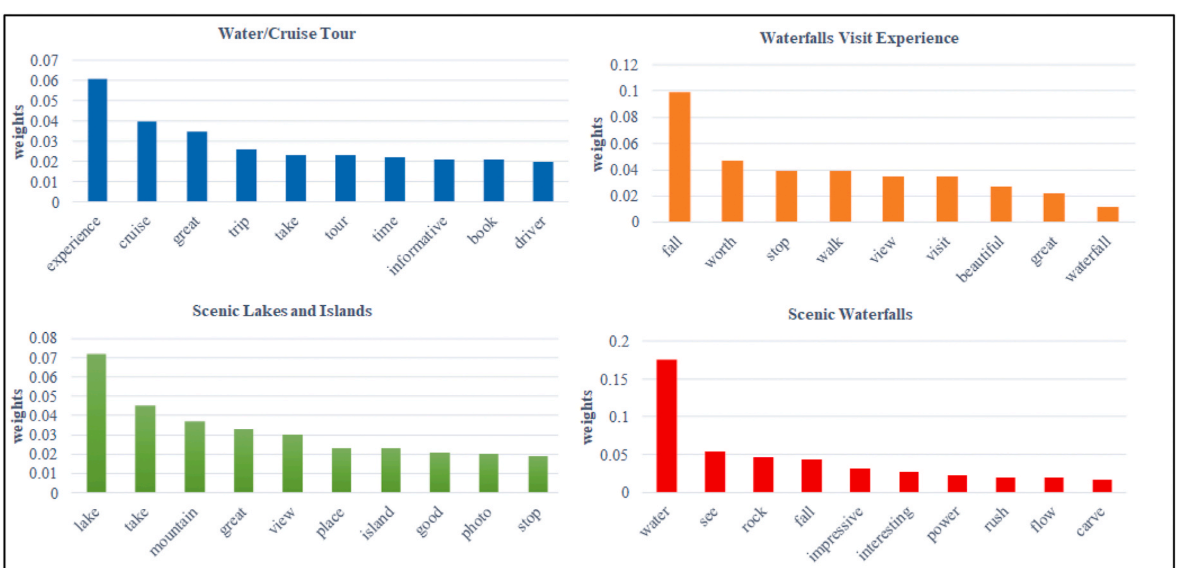

Fig. 4. A sample of 4 topics with relative weights of their top 10 words.

which is based on the top weighted words 'fall,' (99%), 'worth' (47%), 'stop' (39%), 'walk' (39%), 'view' (38%), 'visit' (35%), 'well' (27%), and 'beautiful' (27%) (see Figs. 4 and 5).

The candidate topic label was further tested via rational link to other terms in the top-20 distribution list. If a logical connection was found, the topic name was kept, otherwise, the labeling process restarted using this information to update it. Fig. 6 presents LDA extracted dimensions (topics) from 17,224 online reviews for top touristic locations across JNP. First five dominant topics were 'Ice Walking,' 'Glacier Exploring Tour,' 'Scenic Waterfalls,' 'Water-based Activities,' and 'Waterfalls Visit Experience,' respectively. Three of the dimensions represent tourists' perceptions of glaciers: 'Ice Walking,' 'Glacier Exploring Tour,' and 'Glacier Visit Experience,' while three dimensions correspond to tourists' hiking activities: 'Trails and Pathways,' 'Hiking Activities,' and 'Forest Challenge' (see Table 2). Other groups of dimensions represent lakes and islands (e.g. 'Scenic Lakes and Islands,' 'Water/Cruise Tours,' and 'Water-based Activities') and waterfalls (e.g. 'Scenic Waterfalls' and 'Waterfalls Visit Experience'). The remaining dimensions show four distinct aspects of tourists' general experience (e.g. 'Suggestions,' 'Weather,' 'General Experience'). Fig. 7 also demonstrates Termite two-dimensional visualization of topic models, a visual analysis tool for the term-topic distributions produced by topic models (Chuang, Manning, & Heer, 2012). Termite uses a tabular layout to promote comparison of terms both within and across latent topics and aims to support the domain-specific task of building and refining topic models (Chuang et al., 2012). In Fig. 7, the red bars in the Termite topic model are defined as estimated term frequency within the selected topic, which is equivalence with the relative weights of top words within each topic.

These dimensions were further organized into three fundamental categories based on the Crouch and Ritchie's (1999) model of Tourism Destination Competitiveness (TDC): core resources and attractors, destination management, and qualifying and amplifying determinants (see Table 3). The core resources and attractors refer to the main components of destination that inspire potential visitors to choose one destination over another, or in other words, visitors' key motivators for visiting a tourism destination such as scenic waterfalls and lakes, water-based activities, and hiking activities. These factors are partially-controlled aspects of destination that can be somewhat improved through effective management approaches. Destination management category plays the main role in achieving a balance between all other components of TDC from maintaining and enhancing the core resources and attractors to strengthening of the supporting factors and adjusting with restricting constraints (Crouch, 2011). Destination management components are normally recognized as controlled factors and can be substantially improved by DMOs (e.g. glacier exploring and cruise tours). Finally, qualifying and amplifying determinants of the TDC model refers to the factors that can either positively or negatively drastically affect destination competitiveness (Crouch & Ritchie, 1999; Enright & Newton, 2004). These qualifiers and amplifiers can be alternatively called 'situational conditioners' because they impact tourism demand and are mainly considered as uncontrolled factors such as weather and visitors' perceptions of their own experiences (Ritchie & Crouch, 2011). Tourism providers (e.g. operators of Athabasca Glacier tours, Maligne Canyon Ice Walk tour, boat tours and river cruises) and DMOs should place an emphasis on addressing controlled and some partially-controlled dimensions, such as enhancing tourism experience in glacier and cruise tours, improving infrastructure and informative aspects of hiking trails and pathways, and carefully listening and fulfilling tourists' suggestions and recommendation that are shared online.

Table 2
Classifications of extracted topics into common groups.

Glaciers	Hiking and Trails	Lakes and Islands	Waterfalls	General Experiences
Ice walking	Trails and pathways	Scenic lakes	Scenic waterfalls	Weather
Glacier tour	Hiking activities	Water tours	Waterfalls experience	General experience
Glacier experience	Forest challenge	Water-based activities		Suggestions

4.3. Exploring destination loyalty

After a detailed review of the hospitality and tourism loyalty literature by the author, a vocabulary of destination loyalty keywords was developed. This keyword vocabulary was subsequently used to identify and separate loyalty-expressed reviews from the rest of the corpus. A close reading check was applied to ensure the relatedness of the extracted reviews to destination loyalty conversations. Fig. 8 shows a word cloud of top 100 loyalty-expressed terms toward JNP on TripAdvisor.

Keyword clustering approaches including TF-IDF term document matrix and k-means clustering algorithm have been applied on the corpus of destination loyalty online reviews. The elbow method was applied in order to select the optimal number of clusters (Fig. 9). The elbow method runs k-means clustering on the dataset for a range of values for k (e.g. 1–10), and for each value of k computes an average score for all clusters. When the overall metrics for each model are plotted, and after checking at the percentage of variance explained as a function of the number of clusters, it is possible to visually determine the best value for k by looking at the 'elbow' of the line chart (the point of inflection on the curve) for the best value of k (Ketchen & Shook, 1996). Here, 4 clusters were chosen, as other number of clusters would not have provided improved modeling of the data (Fig. 10).

Results of the k-means clustering method suggest that destination loyalty expressions can be categorized into 4 main subjects, namely

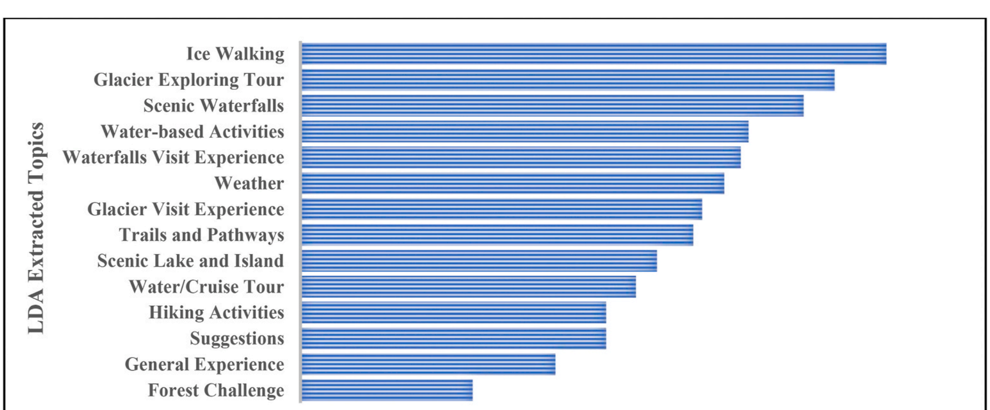

Fig. 6. Extracted topics from LDA topic modeling ranked based on average weights of topic's top words.

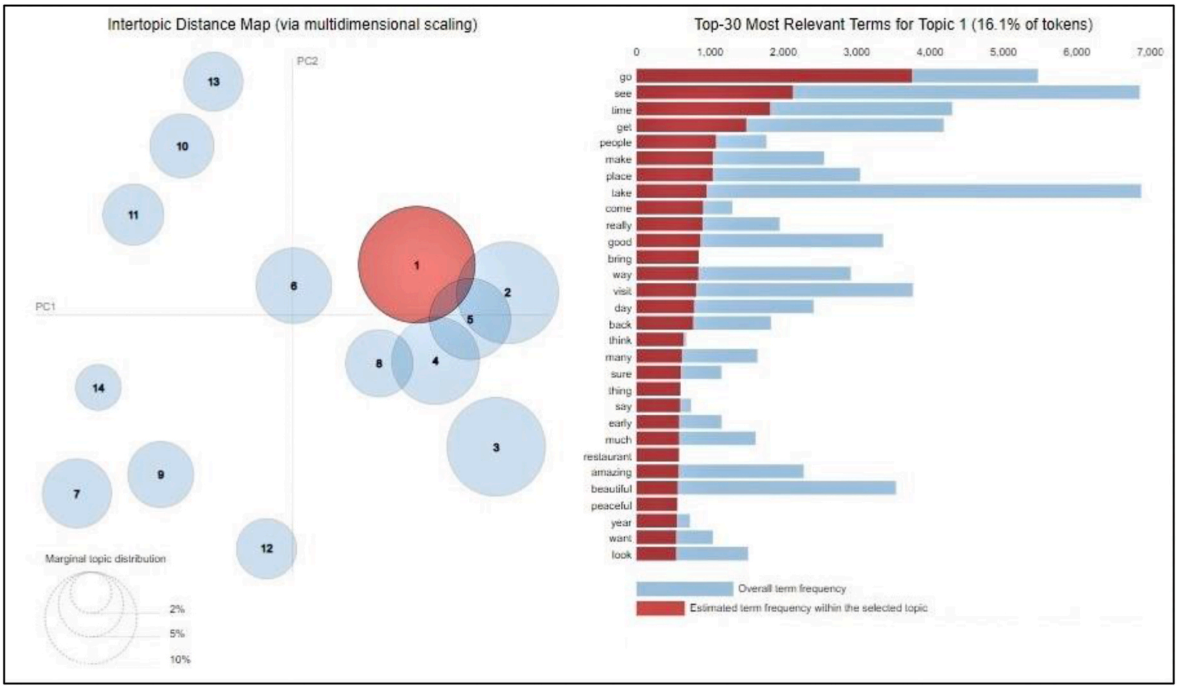

Fig. 7. Two-dimensional visualization of top 30 most relevant terms for the topic 'General Experience' (Termite topic model visualization (Chuang et al., 2012)).

Table 3
A typology of extracted dimensions based on TDC model.

Destination management (controlled)	Core resources and attractors (partially controlled)	Qualifying and amplifying determinants (uncontrolled)
Glacier Exploring Tour	Scenic Lake and Islands	Waterfalls Visit Experience
Water/Cruise Tour	Scenic Waterfalls	Glacier Visit Experience
Trails and Pathways	Water-based Activities	General Experience
Suggestions	Hiking Activities	Weather
	Ice Walking	
	Forest Challenge	

'waterfalls,' 'glaciers,' 'lakes and Islands,' and 'hiking and trails.' This is also aligned with the detected topics of LDA model that were classified into common groups (Table 2). Thus, all of the 12 touristic locations were further categorized and labeled into these 4 clusters based on the nature of the place and types of activities that take place in each (e.g. Athabasca Falls into waterfalls, Athabasca Glaciers into glaciers, Annette Lake into lakes and Islands, and Sulphur Skyline Trails into hiking and trails). Destination loyalty expression reviews with pre-defined labeled clusters were then prepared for a more sophisticated supervised learning clustering (Fig. 11).

Top unigrams and bigrams of each category were identified (Tables 4 and 5). In computational linguistics, N-gram is a contiguous sequence of N items from a given sample of text, therefore, an unigram is referred to a single token (e.g. glacier), and a bigram is a two-word sequence of words (e.g. glacier walk). Bigrams are expected to improve the model performance by taking into consideration words that tend to appear together in the reviews associated with the 12 different locations. A thorough investigation of unigrams and bigrams within each category revealed prevalent characteristics and factors important to tourists when

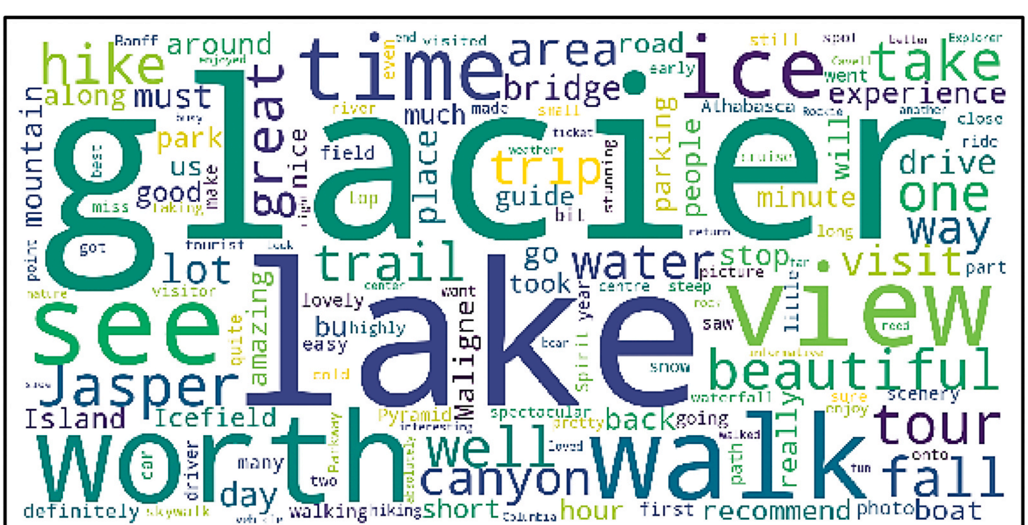

Fig. 8. Word cloud of destination loyalty expressions.

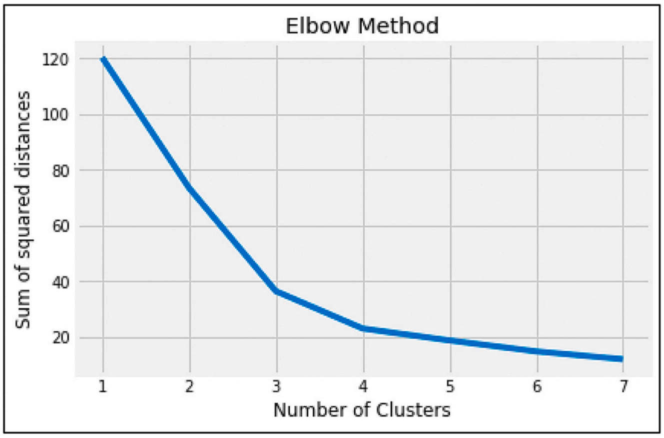

Fig. 9. Elbow method for optimal number of clusters.

expressing their loyalty on SM. In cluster 1, Glaciers, a cross N-gram comparison suggests that tourists' recommendations and revisit intensions were mainly motivated by their experience from Columbia Icefield Glacier Skywalk, entertaining and informative aspects of Ice Explorer Glacier and Icefield Sightseeing tours, and guided interpretive hikes on the Athabasca Glacier (Ice walk). The 230-km mountain road to the Icefields Parkway was also articulated as a worthwhile and beautiful driving experience with spectacular mountain views. In cluster 2, Waterfalls, natural wonders and beautiful sceneries, short walks, less challenging hikes, and winter walks were among the most frequently cited topics of interest for JNP visitors. Cross N-gram comparison within the Lakes and Islands cluster reveals that lake cruise tours, water-based leisure and sport activities (e.g. canoeing/kayaking and fishing), nature and landscape photography, and wildlife viewing are amongst top motivators for visitors to recommend to others and revisit attractions. Finally, the most important aspects in the Hiking and Trails cluster were challenging trails, sense of accomplishment, and beautiful skyline. These dimensions are summarized in Table 6.

4.3.1. *Performance measurement of clustering task*

Linear Support Vector Classifier (also known as Linear SVC) was selected over other classification models (e.g. Random Forest, NB, and Logistic Regression Classifiers) because of a higher accuracy score (Look for Linear SVC in Fig. 12). After classifying and fitting the model to training and test data, the performance of clustering task was evaluated on the test data (i.e. Precision, Recall, and F1-score). The confusion matrix and the classification report of the prediction model for each cluster are described in Fig. 13, where most of the clusters (except for trail most likely due to a smaller sample) have acceptable values of 60% and above.

5. Discussion and conclusions

As SM has been widely adopted by tourists and it has become vital for destinations to leverage their SM platforms to stay competitive in the global economy. SMA is an invaluable method to monitor and listen to consumer-to-consumer conversations (i.e. eWOM) and systematically evaluate tourists' opinions about destinations. Sentiment analysis empowers tourism destinations to track tourists' opinions and viewpoints on a large scale and picture a trajectory of the public buzz around a destination by comparing changes in scores over time and against other places. Destination marketers can also use sentiment analysis to improve customer relationship management and recommendation systems through detecting positive and negative customer feedback. In particular, 'flames' (overly heated or antagonistic languages) can be detected and excluded in social communication to enhance antispam systems (Cambria, Schuller, Xia, & Havasi, 2013).

In this study, the sentiment analysis revealed that some touristic locations in JNP are outperforming others in terms of sentiment and ratio scores on SM, despite the fact that tourists less frequently reflect on their experiences at these places on SM, resulting in lower volumes of reviews (e.g. Sulphur Skyline Trail, Mt. Edith Cavell Trail, and Annette Lake). The presence of these less considered locations placed higher in the ranking suggests that average sentiment score can be a more informative measure than simple TripAdvisor rankings. While the average sentiment score and the Pos/Neg ratio score ranks were aligned with one

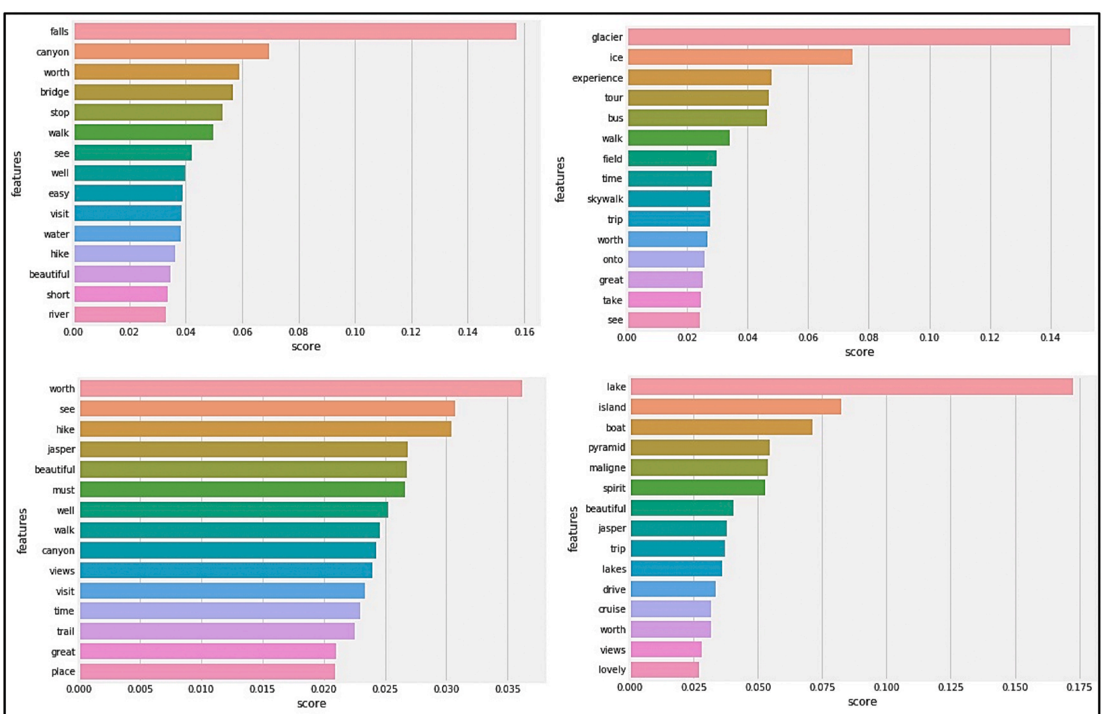

Fig. 10. Features and scores within 4 clusters of destination loyalty expressions.

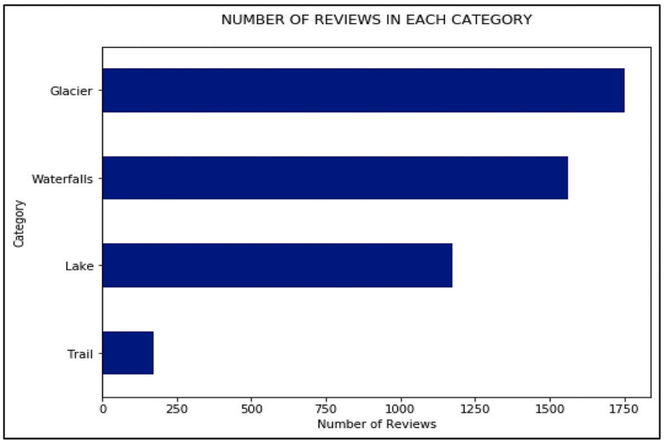

Fig. 11. Distribution of loyalty expression reviews within 4 categories.

Table 4
Top N unigrams between 4 clusters.

Glaciers	Waterfalls	Lakes and Islands	Hiking and Trails
Glacier	Canyon	Lakes	Trail
Skywalk	Waterfalls	Boat	Summit
Icefield	River	Island	Ascent
Experience	Bridges	Cruise	Skyline
Bus	Nature	Quiet	Tough
Tour	Walks	Elk	Hike
Doubt	Gorgeous	Wildlife	Challenging
Winding	Easy	Canoeing	Miette
Icebergs	Paths	Picnic	Rewarded
Cold	Cleats	Beach	Mountain
Entertaining	Routes	Picturesque	Windy
Funny	Amaze	Photography	Climb
Money	Stop	Kayaking	Peak
Informative	Miss	Fishing	Uphill
Drive	Slippery	Moose	Autumn
Considering	Power	Lakeside	Alpine

Table 5
Top N bigrams between 4 clusters.

Glaciers	Waterfalls	Lakes and Islands	Hiking and Trails
ice field	make stop	beautiful lake	hot springs
glacier walk	worth stopping	boat tour	glacial lake
beautiful drive	black bears	lake great	glacier lake
view mountain	easy walk	cruise worth	worth view
worth driving	drive lake	visit lake	amazing scenery
short easy	trip lake	capped mountains	tree line
spectacular view	worth time	pyramid mountain	start trail
fantastic place	beautiful walk	love visit	beautiful glacier
ice worth	ice cleats	love area	beautiful trail
parkway beautiful	hike canyon	tour lake	road little
trails walk	walk view	second visit	viewing point
narrow winding	canyon lake	snow capped	went twice
mountain lake	breathtaking worth	photo stop	miette hot
miss drive	visit way	enjoying view	amazing view
worth experience	visiting jasper	lake frozen	glacier snow
viewing area	easy worth	little island	relatively easy
views drive	early beat	jasper worth	longer hike
beautiful sights	falls beautiful	stop photos	experience time
staff really	beautiful hike	wonderful place	easy moderate
photos worth	special trip	unfortunately weather	saw bear

Table 6
Suggested destination loyalty dimensions within 4 clusters through cross N-gram comparison.

Glaciers	Waterfalls	Lakes & Islands	Hiking and Trails
- Icefield Skywalk	- Natural wonder	- Cruise tours	- Challenging trail
- Ice Explorer tour	- Beautiful scenery	- Water-based activity	- Sense of accomplishment
- Icefield Sightseeing tour	- Short/Easy hike	- Nature photography	- Beautiful skyline
- Icefields Parkway mountain drive	- Winter walk	- Wildlife viewing	

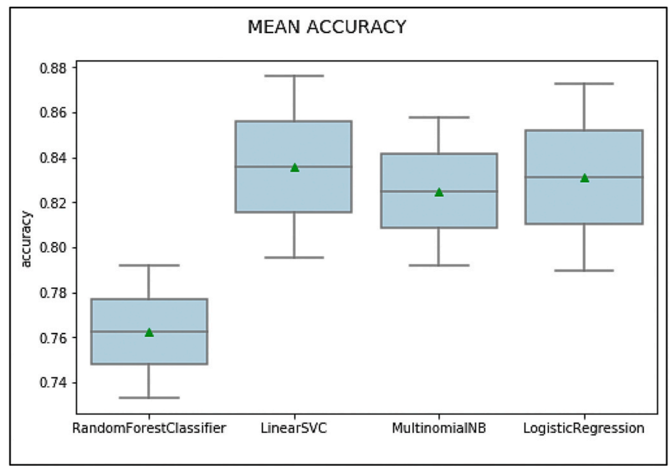

Fig. 12. Accuracy comparison between classification models.

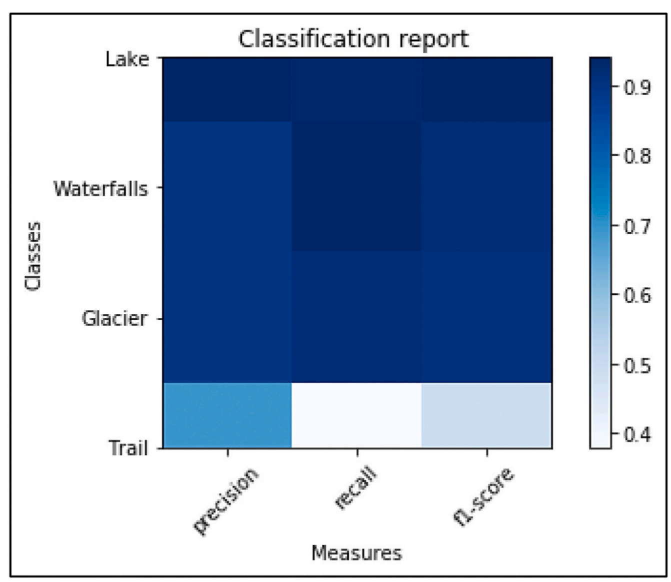

Fig. 13. Accuracy confusion matrix of clustering task.

another for most of the locations, a more detailed review of the scores reveals that some attractions had meaningfully different ranks such as Sulphur Skyline and Mt. Edith Cavell Trails, Spirit Island, and Athabasca Falls. Part of this difference in ranking can be explained by the fact that a higher number of neutral reviews with sentiment scores of zero reduce a location's average score but has no effect on the Pos/Neg ratio score. This suggests that locations with a considerably higher on-average scores compared with their ratio scores may have subgroups of visitors with extremely strong feelings toward these locations (e.g. Sulphur

Skyline and Mt. Edith Cavell Trails).

Sunwapta Falls and Canyon Athabasca Glacier, and Columbia Icefield are located on the other end of sentiment spectrum with the lowest sentiment and ratio scores as well as TripAdvisor ranking. This, however, conflicts with the topic modeling results that suggests glacier activities and tourism (e.g. Ice Walking, Glacier Explore Tours) are amongst the most important dimensions of tourist experience in JNP. Part of this difference can be explained by the fact that conversations around climate change and glacier retreat contain mainly negative expressions and therefore score less in sentiment and ratio rankings. It is quite probable that TripAdvisor follows the same ranking method by relying on sentiment scores, however, this by no means implies that these glacier tourism attractions are less popular in the eyes of JNP tourists.

This study also proposes a novel approach to extract latent dimensions of tourist experience toward a nature-based tourism destination retrieved from online reviews. LDA analysis of online reviews uncovers key aspects that are not discovered through traditional methods. The relative significance of the obtained dimensions is identified based on the intensity of the conversations around each. 'Ice Walking,' 'Glacier Exploring Tour,' 'Scenic Waterfalls,' and 'Water-based Activities' are the most important dimensions in the analysis. This supports the findings of prior studies that have proposed natural environment, beauty of the scenery, and glacier tours as key factors influencing tourism experience and destination image (Beerli & Martin, 2004:; Purdie, 2013). Results of LDA model strongly suggest that JNP tourism providers leverage destination management dimensions (controlled factors) such as glacier and cruise tourism experiences. The quality of the interpretation provided by tour operators and improvement of both content and delivery techniques are crucial factors to optimize tourists' tour experiences. Well planned interpretation more likely results in satisfying visit experiences for tourists, which in turn leads to positive word-of-mouth, recommendations, and repeat visitation (Hwang, Lee, & Chen, 2005). DMOs and tour operators can play an important role in filling the knowledge gaps through trainings and workshops, mentoring and internships, as well as providing information materials directly to tourists pre, during and post visit. This goal cannot be achieved without a clear and effective communication and liaison channels between DMOs and tour operators.

Another controlled dimension was trails and pathways. Considering the exceptionally higher sentiment scores of trails and hiking locations, both from our results and on TripAdvisor rankings, improving infrastructure and informative aspects of hiking trails and pathways through strategic and operational plans for trail development is something that tourism providers should invest further in. DMOs should also understand the needs and characteristics of potential hikers, identify diverse constraints that prevent their trail use, and recognize factors that inspire and facilitate their use. DMOs can also develop partnerships across different public and private sectors to promote specific trail activities, hiking experiences and packages for target groups, for example through showcasing unique cultural, natural, and historical features of the trail.

This study advances investigations of destination loyalty through cluster analysis of TripAdvisor online reviews. A destination loyalty keyword vocabulary was developed through reviewing loyalty literature in hospitality and tourism, and loyalty-expressed reviews were identified and separated from the rest of the JNP TripAdvisor corpus. After categorizing loyalty-expressed reviews into 4 clusters of glaciers, waterfalls, lakes and islands, and hiking and trails, top features within each cluster were presented and analyzed. Results revealed that different types of tours play an important role in recommendations and revisit intentions of JNP tourists (e.g. Columbia Icefield and Sightseeing tours, Glacier Skywalk, Maligne Lake Cruise tour). Water-based recreational activities such as kayaking and canoeing, boating, paddle boarding, and fishing were amongst highly recommended activities when visiting lakes and islands. Nature photography and wildlife viewing were other inspiring factors for destination loyalty expression in reviews. Aligned with the findings from sentiment analysis and topic modeling, hiking activities and trail attractions were notable motivators for tourists' loyalty expressions on SM. Results show that sense of accomplishment upon finishing longer hikes and more challenging trails together with beautiful skyline and alpine view are amongst reasons for sharing loyalty toward JNP online.

This study has several managerial implications. Tourism providers can not only verify underlying aspects of tourist experience from user-generated data but can also portray a perceptual mapping of touristic locations within their destination through a comprehensive analysis of online reviews. Moreover, there is a lack of understanding about the factors influencing destination loyalty in nature-based setting. Thus, this study enables DMOs to specify destination's salient characteristics that influence tourists' recommendations and revisits intentions. Online review analysis of JNP visitors reveals key dimensions of destination loyalty toward JNP, including informative and recreational tours, water-based recreational activities, and challenging trails.

While the findings of this study contribute to the academia and tourism industry, it has some limitations. First and foremost, due to the inherently incomplete nature of online reviews (Chen et al., 2016; Hargittai, 2020; Morstatter & Liu, 2017), SM analyses must be treated with caution, especially when applying and comparing results beyond the particular online groups under study. Combining SM data with other types of data such as interviews, focus groups, and surveys may be a useful strategy for tourism researchers to not only distinguish extreme views from more typical perspectives held by tourists but to generalize their results to other populations. Second, it is hard to generalize the findings to other tourism destinations due to the exploratory nature of this study. Thus, future research can replicate the current study in other destinations to test the applicability of data analysis and compare the findings from attractions and tourism destinations across the globe. Another limitation of the current study is the comprehensiveness of the collected data from different touristic locations within JNP, as well as focusing only on TripAdvisor. Future research can not only make use of a broader scope and include other touristic places but can also incorporate other SM sources such as Twitter to better understand tourists' sentiments and interests. Last but not least, the current study treated the entire extracted reviews from 2002 to 2019 as one corpus, and failed to analyze the trend components of the time-series data. Future research is encouraged to divide and compare different time spans in the data and explore how tourists' behaviours and attitudes change over time.

Declaration of competing interest

None.

APPENDICES

Appendix A

Latent dirichlet allocation (LDA) topic modeling (Blei et al., 2003) is a generative statistical model used to find hidden semantic structures of textual content and is helpful for discovering the main topics and meaningful dimensions of online reviews. LDA model assumes that a set of topics and themes exists in the text and tries to uncover these hidden structures by looking at the co-occurrence of content terms in the text. In other words, LDA model repeatedly samples the words of the corpus based on a multinomial distribution to identify words that tend to associate with each other. The outputs of LDA model are the list of topics, surfaced based on the likelihood of word co-occurrence, and weight values presenting the probability that a word belongs to a specific topic. Topic models based on LDA technique are frequently used as a text-mining method to discover the hidden semantic structures in a text; however, evaluating the intrinsic quality of the topic model and topics remains controversial.

Appendix B

One of the very first methodologies for evaluating the semantic interpretability of topics was introduced by Chang et al. (2009) as a 'word intrusion' indirect approach, where 'intruder words' are randomly inserted into LDA output topics and human annotators try to identify the intruded words. Newman et al. (2010) introduced the notion of 'topic coherence' and tried to estimate the human-interpretability of topics using a more direct approach. In this method, human annotators were asked to rate topics on a three-point scale based on the coherence level of the topic words. They then assessed topic coherence based on pairwise pointwise mutual information (PMI) between the topic words. One of the biggest limitations of these methods is that they underperform in large-scale evaluations since they require human annotations (Lau et al., 2014). Lau et al. (2014) introduced an improved formulation of Newman et al.'s (2010) approach based on normalized PMI (NPMI), a fully automated word intrusion method (WI-Auto-NPMI) and observed coherence (OC-Auto-NPMI) tasks. Their results show that NPMI achieves a noticeably higher correlation than OC-Human, especially at the model level.

Appendix C

With the assumption that a given online review is subjective, sentiment analysis represents a polarity classification and valence identification of reviews and determines whether the polarity of textual content is positive, negative, or neutral. In the tourism context, this polarity classification of 'positive' and 'negative' can be inferred as 'satisfied' and 'dissatisfied,' respectively (Alaei et al., 2019). The lexicon-based approach of sentiment analysis compares tokens of a given online review to pre-defined positive and negative sentiment lexicons to determine whether the review has a more positive or negative tone. In a supervised method of classification, a training dataset is first developed to distinguish a document's characteristics, and is further applied to test data (Feldman, 2013).

Appendix D

One important step toward an effective text clustering process is that word frequencies need to be normalized in terms of their relative frequency of occurrence in the document and over the entire corpus. This task can be performed by vector-space based TF-IDF representation, where the TF for each word is normalized by the Inverse Document Frequency (IDF). The IDF normalization reduces the weight of more frequent terms in the corpus (e.g. stop-words), ensuring that the matching of documents is more influenced by unique words with relatively low frequencies. A sub-linear transformation function is also normally applied to the term frequencies in order to avoid the adverse effects of having a single term that might be very frequent in a document (Aggarwal & Zhai, 2012).

Credit roles

Farshid Mirzaalian: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization: Elizabeth Halpenny: Conceptualization, Methodology, Validation, Writing - Reviewing and Editing, Supervision.

References

Aggarwal, C. C., & Zhai, C. (2012). A survey of text clustering algorithms. In *Mining text data* (pp. 77–128). Boston, MA: Springer.

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research, 58*(2), 175–191.

Alexandris, K., Kouthouris, C., & Meligdis, A. (2006). Increasing customers' loyalty in a skiing resort. *International Contemporary Hospitality Management, 18*(5), 414–425.

Backman, S. J., & Crompton, J. L. (1991). Differentiating between high, spurious, latent, and low loyalty participants in two leisure activities. *Journal of Park and Recreation Administration, 9*(2), 1–17.

Baker, D. A., & Crompton, J. L. (2000). Quality, satisfaction and behavioral intentions. *Annals of Tourism Research, 27*(3), 785–804.

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

Berezan, O., Raab, C., Tanford, S., & Kim, Y. S. (2015). Evaluating loyalty constructs among hotel reward program members using eWOM. *Journal of Hospitality & Tourism Research, 39*(2), 198–224.

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

Bolton, R. N., & Lemon, K. N. (1999). A dynamic model of customers' usage of services: Usage as an antecedent and consequence of satisfaction. *Journal of Marketing Research, 36*(2), 171–186.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210–230.

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the internet—the state of eTourism research. *Tourism Management, 29*(4), 609–623.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems, 28*(2), 15–21.

Capriello, A., Mason, P. R., Davis, B., & Crotts, J. C. (2013). Farm tourism experiences in travel reviews: A cross-comparison of three alternative methods for data analysis. *Journal of Business Research, 66*(6), 778–785.

Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., & Blei, D. M. (2009, December). Reading tea leaves: How humans interpret topic models. In *Neural information processing systems* (Vol. 22), 288-296.

Chen, C. F., & Chen, F. S. (2010). Experience quality, perceived value, satisfaction and behavioral intentions for heritage tourists. *Tourism Management, 31*(1), 29–35.

Chen, J. S., & Gursoy, D. (2001). An investigation of tourists' destination loyalty and preferences. *International Journal of Contemporary Hospitality Management, 13*(2), 79–85.

Chen, C. F., & Tsai, D. (2007). How destination image and evaluative factors affect behavioral intentions? *Tourism Management, 28*(4), 1115–1122.

Chen, H., Zheng, Z., & Ceran, Y. (2016). De-biasing the reporting bias in social media analytics. *Production and Operations Management, 25*(5), 849–865.

Chi, C. G. Q., & Qu, H. (2008). Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach. *Tourism Management, 29*(4), 624–636.

Chiu, C., Chiu, N. H., Sung, R. J., & Hsieh, P. Y. (2015). Opinion mining of hotel customer-generated contents in Chinese weblogs. *Current Issues in Tourism, 18*(5), 477–495.

Chuang, J., Manning, C. D., & Heer, J. (2012). May). Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the international working conference on advanced visual interfaces* (pp. 74–77).

Chua, A., Servillo, L., Marcheggiani, E., & Moere, A. V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management, 57*, 295–310.

Crouch, G. I. (2011). Destination competitiveness: An analysis of determinant attributes. *Journal of Travel Research, 50*(1), 27–45.

Crouch, G. I., & Ritchie, J. B. (1999). Tourism, competitiveness, and societal prosperity. *Journal of Business Research, 44*(3), 137–152.

Del Vecchio, P., Mele, G., Ndou, V., & Secundo, G. (2018). Creating value from social big data: Implications for smart tourism destinations. *Information Processing & Management, 54*(5), 847–860.

Enright, M. J., & Newton, J. (2004). Tourism destination competitiveness: A quantitative approach. *Tourism Management, 25*(6), 777–788.

Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management, 52*, 498–506.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM, 56*(4), 82–89.

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.

Gallarza, M. G., & Saura, I. G. (2006). Value dimensions, perceived value, satisfaction and loyalty: An investigation of university students' travel behaviour. *Tourism Management, 27*(3), 437–452.

Gavcar, E., & Gursoy, D. (2002). Research note: An examination of destination-originated (pull) factors. *Tourism Analysis, 7*(1), 75–81.

George, B. P., & George, B. P. (2004). Past visit and the intention to revisit a destination: Place attachment as the mediator and novelty seeking as the moderator. *Journal of Tourism Studies, 15*(2), 51–66.

Gilbert, C. H. E., & Hutto, E. (2014, June). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *, Vol. 81. Eighth international conference on weblogs and social media (ICWSM-14)* (p. 82). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf.

Gilbert, G. R., & Veloutsou, C. (2006). A cross-industry comparison of customer satisfaction. *Journal of Services Marketing, 20*(5), 298–308.

Gligorijevic, B. (2016). Review platforms in destinations and hospitality. In R. Egger, I. Gula, & D. Walcher (Eds.), *Open tourism* (pp. 215–228). Berlin, Heidelberg: Springer.

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.

Gursoy, D., S Chen, J., & G Chi, C. (2014). Theoretical examination of destination loyalty formation. *International Journal of Contemporary Hospitality Management, 26*(5), 809–827.
Halpenny, E. A. (2006). *Environmental behaviour, place attachment and park visitation: A case study of visitors to point pelee national park [unpublished doctoral dissertation]*. University of Waterloo.
Halpenny, E. A., Kulczycki, C., & Moghimehfar, F. (2016). Factors effecting destination and event loyalty: Examining the sustainability of a recurrent small-scale running event at Banff National Park. *Journal of Sport & Tourism, 20*(3–4), 233–262.
Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media. *Social Science Computer Review, 38*(1), 10–24.
Hays, S., Page, S. J., & Buhalis, D. (2013). Social media as a destination marketing tool: Its use by national tourism organisations. *Current Issues in Tourism, 16*(3), 211–239.
Hofmann, T. (1999, August). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 50–57).
Huang, A. (2008, April). Similarity measures for text document clustering. In *Proceedings of the sixth New Zealand computer science research student conference (NZCSRSC2008)* (Vol. 4, pp. 9–56). Christchurch: New Zealand.
Hwang, S. N., Lee, C., & Chen, H. J. (2005). The relationship among tourists' involvement, place attachment and interpretation satisfaction in Taiwan's national parks. *Tourism Management, 26*(2), 143–156.
Iso-Ahola, S. E. (1982). Toward a social psychological theory of tourism motivation: A rejoinder. *Annals of Tourism Research, 9*(2), 256–262.
Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal, 17*(6), 441–458.
Kim, K., Hallab, Z., & Kim, J. N. (2012). The moderating effect of travel experience in a destination on the relationship between the destination image and the intention to revisit. *Journal of Hospitality Marketing & Management, 21*(5), 486–505.
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.
Költringer, C., & Dickinger, A. (2015). Analyzing destination branding and image from online sources: A web content mining approach. *Journal of Business Research, 68*(9), 1836–1843.
Kozinets, R. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research, 39*, 61–72.
Kyle, G., Graefe, A., Manning, R., & Bacon, J. (2004). Predictors of behavioral loyalty among hikers along the Appalachian Trail. *Leisure Sciences, 26*(1), 99–118.
Lau, J. H., Newman, D., & Baldwin, T. (2014). April). Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In *Proceedings of the 14th conference of the European chapter of the association for computational linguistics* (pp. 530–539).
Lee, J., Graefe, A. R., & Burns, R. C. (2007a). Examining the antecedents of destination loyalty in a forest setting. *Leisure Sciences, 29*(5), 463–481.
Lee, C. K., Yoon, Y. S., & Lee, S. K. (2007b). Investigating the relationships among perceived value, satisfaction, and recommendations: The case of the Korean DMZ. *Tourism Management, 28*(1), 204–214.
Li, Y. R., Lin, Y. C., Tsai, P. H., & Wang, Y. Y. (2015). Traveller-generated contents for destination image formation: Mainland China travellers to Taiwan as a case study. *Journal of Travel & Tourism Marketing, 32*(5), 518–533.
Lu, W., & Stepchenkova, S. (2012). Ecotourism experiences reported online: Classification of satisfaction attributes. *Tourism Management, 33*(3), 702–712.
Mason, M. C., & Nassivera, F. (2013). A conceptualization of the relationships between quality, satisfaction, behavioral intention, and awareness of a festival. *Journal of Hospitality Marketing & Management, 22*(2), 162–182.
Melbourne, Australia McKay, K., Van Winkle, C., & Halpenny, E. (2019). Active vs passive social media use, attendee engagement, and festival loyalty. In *Presented at 2019 travel and tourism research association: Advancing tourism research globally*. June 2019 https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=2263&context=ttra.
Menner, T., Höpken, W., Fuchs, M., & Lexhagen, M. (2016). Topic detection: Identifying relevant topics in tourism reviews. In A. Inversini, & R. Schegg (Eds.), *Information and communication technologies in tourism 2016* (pp. 411–423). Berlin, Heidelberg: Springer.
Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A big data analytics method for tourist behaviour analysis. *Information & Management, 54*(6), 771–785.
Milman, A., & Pizam, A. (1995). The role of awareness and familiarity with a destination: The central Florida case. *Journal of Travel Research, 33*(3), 21–27.
Miner, G., Elder IV, J., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Netherland, Amsterdam: The Academic Press.
Mirzaalian, F., & Halpenny, E. (2019). Social media analytics in hospitality and tourism: A systematic literature review and future trends. *Journal of Hospitality and Tourism Technology, 10*(4), 764–790.
Morais, D. B., Dorsch, M. J., & Backman, S. J. (2004). Can tourism providers buy their customers' loyalty? Examining the influence of customer-provider investments on loyalty. *Journal of Travel Research, 42*(3), 235–243.
Morstatter, F., & Liu, H. (2017). Discovering, assessing, and mitigating data bias in social media. *Online Social Networks and Media, 1*, 1–13.
Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). June). Automatic evaluation of topic coherence. In *Human language technologies: The 2010 annual conference of the north American chapter of the association for computational linguistics* (pp. 100–108).
Octoparse. (2019). *About Octoparse*. Retrieved 11 November 2019. From: https://www.octoparse.com/.
Oppermann, M. (2000). Tourism destination loyalty. *Journal of Travel Research, 39*(1), 78–84.
Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of empirical methods in natural language processing* (pp. 79–86).
Pantano, E., Priporas, C. V., & Stylos, N. (2017). 'You will like it!' Using open data to predict tourists' response to a tourist attraction. *Tourism Management, 60*, 430–438.
Parks Canada. (2019). *About jasper national park*. Retrieved 11 November 2019. From: https://www.pc.gc.ca/en/pn-np/ab/jasper/.
Pearce, P. L., & Wu, M. Y. (2018). Entertaining international tourists: An empirical study of an iconic site in China. *Journal of Hospitality & Tourism Research, 42*(5), 772–792.
Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management, 55*, 16–24.
Prayag, G., & Ryan, C. (2012). Antecedents of tourists' loyalty to Mauritius: The role and influence of destination image, place attachment, personal involvement, and satisfaction. *Journal of Travel Research, 51*(3), 342–356.
Prebensen, N. K., Woo, E., Chen, J. S., & Uysal, M. (2013). Motivation and involvement as antecedents of the perceived value of the destination experience. *Journal of Travel Research, 52*(2), 253–264.
Pritchard, M. P., & Howard, D. R. (1997). The loyal traveler: Examining a typology of service patronage. *Journal of Travel Research, 35*(4), 2–10.
Purdie, H. (2013). Glacier retreat and tourism: Insights from New Zealand. *Mountain Research and Development, 33*(4), 463–473.
Reid, E. (1996). Informed consent in the study of on-line communities: A reflection on the effects of computer-mediated social research. *The Information Society, 12*, 119–127.
Ritchie, J. B., & Crouch, G. I. (2011). A model of destination competitiveness and sustainability. In Y. Wang, & A. Pizam (Eds.), *Tourism destination marketing and management: Collaborative strategies* (pp. 326–357). Wallingford: CABI Publishing.
Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing, 32*(5), 608–621.
Senders, A., Govers, R., & Neuts, B. (2013). Social media affecting tour operators' customer loyalty. *Journal of Travel & Tourism Marketing, 30*(1–2), 41–57.
Sönmez, S. F., & Graefe, A. R. (1998). Determining future travel behavior from past travel experience and perceptions of risk and safety. *Journal of Travel Research, 37*(2), 171–177.
Sparks, B. (2007). Planning a wine tourism vacation? Factors that help to predict tourist behavioural intentions. *Tourism Management, 28*(5), 1180–1192.
Statista.com. (2019). *TripAdvisor - statistics & facts*. Retrieved 11 November 2019. From: https://www.statista.com/topics/3443/tripadvisor/.
Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: A social media analytics framework. *Social Network Analysis and Mining, 3*(4), 1277–1291.
Sun, X., Chi, C. G. Q., & Xu, H. (2013). Developing destination loyalty: The case of Hainan Island. *Annals of Tourism Research, 43*, 547–577.
Tian-Cole, S., & Crompton, J. (2003). A conceptualization of the relationships between service quality and visitor satisfaction, and their links to destination selection. *Leisure Studies, 22*(1), 65–80.
Um, S., Chon, K., & Ro, Y. (2006). Antecedents of revisit intention. *Annals of Tourism Research, 33*(4), 1141–1158.
Velázquez, B. M., Saura, I. G., & Molina, M. E. R. (2011). Conceptualizing and measuring loyalty: Towards a conceptual model of tourist loyalty antecedents. *Journal of Vacation Marketing, 17*(1), 65–81.
Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management, 46*, 222–232.
Wartena, C., & Brussee, R. (2008, September). Topic detection by clustering keywords. In *2008 19th international workshop on database and expert systems applications* (pp. 54–58).
Williams, D. R., & Roggenbuck, J. W. (1989, October). Measuring place attachment: Some preliminary results. In *Paper presented at the session on outdoor planning and management, NRPS symposium on leisure research* (San Antonio, Texas).
Witt, C. A., & Wright, P. (1992). Tourist motivation: Life after maslow. In P. Johnson, & B. Thomas (Eds.), *Choice and demand in tourism* (pp. 33–56). London: Mansell.
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., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management, 31*(2), 179–188.
Xiang, Z., Schwartz, Z., Gerdes, J. H., Jr., & 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.
Xie, K. L., So, K. K. F., & Wang, W. (2017). Joint effects of management responses and online reviews on hotel financial performance: A data-analytics approach. *International Journal of Hospitality Management, 62*, 101–110.
Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications, 36*(3), 6527–6535.
Yoo, K. H., & Gretzel, U. (2008). What motivates consumers to write online travel reviews? *Information Technology & Tourism, 10*(4), 283–295.
Yoon, Y., & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: A structural model. *Tourism Management, 26*(1), 45–56.
Yuksel, A., Yuksel, F., & Bilim, Y. (2010). Destination attachment: Effects on customer satisfaction and cognitive, affective and conative loyalty. *Tourism Management, 31*(2), 274–284.
Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of Consumer Research, 12*(3), 341–352.

Zeng, D., Chen, H., Lusch, R., & Li, S. H. (2010). Social media analytics and intelligence. *IEEE Intelligent Systems, 25*(6), 13–16.
Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives, 10*, 27–36.
Zhang, H., Fu, X., Cai, L. A., & Lu, L. (2014). Destination image and tourist loyalty: A meta-analysis. *Tourism Management, 40*, 213–223.
Zhou, X., Xu, C., & Kimmons, B. (2015). Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Computers, Environment and Urban Systems, 54*, 144–153.

Farshid Mirzaalian – First author: Farshid Mirzaalian, B.Eng., MBA, PhD, is a researcher in the area of hospitality and tourism marketing in University of Alberta, Canada. Farshid's research interests include consumer behaviour analysis, sentiment analysis and opinion mining, text analytics, social media analytics, spatial analysis, big data and Geographic Information Systems (GIS) and business analytics in the hospitality and tourism.

Elizabeth Halpenny – co-author: A Professor at Canada's University of Alberta, Elizabeth Halpenny, PhD, teaches and conducts research in the areas of tourism, marketing, environmental psychology and protected areas management. Elizabeth's research focuses on individual's interactions with nature environments, tourism experience, and environmental stewardship. Current research projects include: (a) the effect of mobile digital technologies on visitors' experiences: (b) the impact of World Heritage designation and other park-related brands on travel decision making; (c) individuals' attitudes towards and stewardship of natural areas; and (d) children, health and nature.