A big-data approach for investigating destination image gap in Sanya City: When will the online and the offline goes parted?

Lingkun Meng  
*School of Tourism Management, Sun Yat-sen University, China*

Yi Liu  
*School of Tourism Management, Sun Yat-sen University, China*, liuyi89@mail.sysu.edu.cn

Yuanlei Wang  
*School of Tourism Management, Sun Yat-sen University, China*

Xiaojuan Li  
*School of Tourism Management, Sun Yat-sen University, China*

Follow this and additional works at: https://egijournals.researchcommons.org/regional-sustainability

**Recommended Citation**

Meng, Lingkun; Liu, Yi; Wang, Yuanlei; and Li, Xiaojuan (2021) "A big-data approach for investigating destination image gap in Sanya City: When will the online and the offline goes parted?," *Regional Sustainability*  Vol. 2 : Iss. 1 , Article 9.  
DOI: 10.1016/j.regsus.2021.02.001

Available at: https://egijournals.researchcommons.org/regional-sustainability/vol2/iss1/9

This Full Length Article is brought to you for free and open access by Journals of EGI. It has been accepted for inclusion in Regional Sustainability by an authorized editor of Journals of EGI.
A big-data approach for investigating destination image gap in Sanya City: When will the online and the offline goes parted?

Lingkun Meng, Yi Liu*, Yuanlei Wang, Xiaojuan Li

School of Tourism Management, Sun Yat-sen University, Zhuhai, 519082, China

ARTICLE INFO

Keywords:
Big-data
Tourism reviews
Destination image
“Cognitive-affective” model
Tourism sentiment evaluation model

ABSTRACT

Tourism destination images in terms of the gaps between the projected and perceived images are of great significance in the development of destinations. Additionally, the use of big-data in tourism studies remains under-utilized despite the boom in big-data applications and the increasing number of electronic User Generated Contents (UGC). Aiming to take advantage of tourism UGC to fully understand the destination image gap between official promotion materials and tourist perception of Sanya City in China, this study innovatively employed a big-data analysis technique, Tourism Sentiment Evaluation (TSE) model and proposed a new analysis framework integrating the “cognitive-affective” model with the gap analysis of projected and perceived destination image to explore the destination image gap of Sanya. It is found that Sanya’s perceptive destination image is overall consistent with its official positioning; however, there also exist image gaps between the two groups in terms of the impact of festival events and tourists’ attitude towards core scenic spots amongst others. This study’s findings are discussed in light of their methodological, theoretical, and practical implications for destination positioning, marketing, and management.

1. Introduction

Tourism destination image is one of the most important research topics in evaluating tourism destination development in human geography and management science. Good destination images are able to improve tourist satisfaction and tourist revisit intention ratio, and at the same time, transmitting positive social and cultural meanings effectively for destination branding. Nevertheless, in the process of destination branding, there may exist disparities or even contradictions between a destination’s typical characteristics and its eventual image stemming from official publicity materials (Andreu et al., 2000; Grosspietsch, 2006; Zhang and Wu, 2016; Marine-Roig and Ferrer-Rosell, 2018). Consequently, the disparity in the portrayal of the destination images will make it difficult for the destination to meet the expectations of visiting tourists and sustain destination appeal. By conducting a comparative analysis between official publicity images and the tourists’ perception of authenticity and uncovering the disparity between them, this study can contribute towards existing literature on destination positioning and marketing.

The analysis of destination image gaps has traditionally been measured by questionnaires and other general statistical tools. Lately, the emergence and boom in the usage of big-data in tourism studies have contributed to the development of tourism research and related fields. The conventional quantitative methods of questionnaires or web texts have inherent limitations in sample size and such limitations affect the richness and credibility of the data collected and analyzed (Liu et al., 2017a). In contrast, as emerging research materials, electronic tourism reviews have an advantage in terms of quantity and refined content (Zhang and Luo, 2017; Liu et al.,...
2019a). However, despite the potential of big-data applications in data collection resources such as electronic User Generated Contents (UGC), the uses of these methods remain under-utilized in tourism studies despite easy accessibility.

Seeking a departure from the existing quantitative methods, this paper aims to use big-data analysis techniques comprising of lexicon filtering methods and co-occurrence analysis so as to provide a new and innovative approach for measuring the efficacy of destination images. Using a famous tourism destination as the site of analysis namely Sanya City in South China, this study aims to uncover the disparity in perception between official destination images against tourist perception. We believe that this study will enable Sanya to promote its image better and in the process, to attract even more tourists in future. We also propose general suggestions for destinations similar to Sanya.

2. Tourism destination branding and gaps

2.1. Gaps in the process of destination branding

The concept of destination image has aroused great interest in academic circles since it was first proposed in the 1970s (Hunt, 1975). Among these studies, there are accomplished works focused on unearthing the implications of a mismatch between ideal and projected images on destination image and tourist satisfaction (Cai, 2002). Given that the destination image is often related to the destination’s future development, comparative studies of the destination image of different groups will, in turn, contribute to advance understanding of destination image and its implications (Andreu et al., 2000). There are also comparative studies about the differences between tourism practitioners and tourists as these two groups of stakeholders are closely connected and play important roles in the development of destinations (Chon, 1990; Andreu et al., 2000; Daniel, 2018). Grosspietsch (2006) postulated that there are two sides to a destination image-perceived image by tourists and projected image by tourism operators. The concept of projected image, referring to the image that tourism operators try to establish, originates from the study of tourism destination marketing (Barich and Kotler, 1991). The projected image is the operator’s idealized image of all the critical elements of the promoted destination. An important purpose of Destination Marketing Organization (DMO) is to attract tourists based on the appeal of attractive destination images (Pike and Page, 2014). Tourist perception is therefore of great significance to DMOs as it enables tourist operators to improve the marketing of their destinations, optimize their management modes, as well as create a deeper experience for visitors (Pike and Ryan, 2004).

However, the conclusions from the comparative studies of DMO and tourists to date have been surprising. It had been proven that the planned image launched by DMO in the tourism market has not been recognized by tourists (Zhang and Wu, 2016). Similar outcomes have been found in many other studies. For example, there have been discussions on the differences between the projected and perceived image of Spain amongst British tourists (Andreu et al., 2000). Song and Kim (2016) discovered the differences in destination image between the official website of DMO and tourists’ social networking sites, by analyzing a large number of images on these two channels of Kyoto, Tokyo, and Osaka, Japan. Grosspietsch (2006) compared the differences in images of Rwanda as a tourist destination between tourists and operators, and realized that there are significant differences in perception between these two groups in terms of the security situation, tourist amenities, and the values of interaction with locals. The gap between projected and perceived destination images of Catalonia, Spain, was also measured by applying compositional analysis (Marine-Roig and Ferrer-Rosell, 2018).

Identification and measurement of the scale and depth of the destination image gap is often a difficult task. It takes substantial resources and continuous efforts to conduct in-site surveys for this endeavor. In the process of destination image building, there are often communication gaps (Camprubi et al., 2008). For cities, these gaps often stem from the mismatch between expectations and diversified perceptions across major city image stakeholders, such as government, tourists, and residents (Zenker et al., 2017). This disparity is the first obstacle in brand communication from the hosts (destination and DMO) to the guests (tourists and outsiders) and in turn, hinders effective city branding and image strategies. To alleviate this issue and to optimize resources in tourism image planning, it is pertinent to examine the differences with the aim of understanding and pinpointing the discrepancies (Klingmann, 2007).

2.2. Formation and perception of tourism destination image

Depending on whether potential tourists have visited the destination yet, a tourism destination image might be composed of organic image and induced image (Gunn, 1988). Shortly after Gunn’s finding (Gunn, 1988), the idea of affective destination image was introduced into the literature and the focus of subsequent analyses in this field was shifted from pure cognition-driven to a combination of cognition and affective studies in destination image studies. To be specific for our purpose, cognitive destination image refers to the subject’s perception of the attributes of a tourism destination, while affective destination image refers to the emotional responses of subjects to the place (Baloglu and Mcleary, 1999). Using a quantitative approach leaned towards questionnaires, Baloglu and Mcleary (1999) concluded that tourists’ cognitive and affective components might directly affect the total destination image, which constitutes the theoretical foundation of the widely used measurements of tourism destination image. Subsequently, Beerli and Martin (2004) proposed to separate a cognitive image into four levels: resources, facilities, atmosphere, and social environment. In affective image analysis, the extant researches are based on the opposite affective ring model proposed by Feldman and Russel (1998), and the main research method is through the analysis of high-frequency adjectives (Andsager and Drzewiecka, 2002). Not just in a perceived image but the projected image also consists of a cognitive and affective evaluation of DMO.
Recently, Feng et al. (2018) took eight national 5A scenic spots in Shaanxi Province, China, as the basis of their case study and used the ROST Content Mining software, dissected the disparity in destination image of Shaanxi tourism destinations from the tourists’ perception and emotional responses through data mining of network travel notes. They also explored the moderating effects of DMO electronic platforms on the destination image through a conceptual model and it is found that tourist involvement has a positive impact on cognitive and affective image, and the impression of the destination image is an antecedent of visitation intention (Molinillo et al., 2018). Moreover, over three key stages in the travel itinerary (before, during, and after travel), the transition of perception in images including affective, cognitive, and overall image are identified to enhance the understanding of how image develops through these three key stages of a trip (Kim et al., 2019).

2.3. Methods for destination image measurement

Studies of projected images often analyze official publicity materials including the official tourism website, guide book, and tourism magazine as their research basis. For instance, promotional materials such as brochures and travel guides issued by the Qingdao Municipal Government and Tourism Administration, Shandong Province of China were used to understand the official projected image of Qingdao (Ji and Wall, 2014). The projected image of a country was investigated using different provinces/cities’ logos (Ngheim-Phu, 2015).

The techniques measuring destination image perception may be divided into two main types: structural and non-structural methods (Yang et al., 2007) (see Table 1). Structural analysis methods select different evaluation factors and use a standard model to obtain the image perception, through the collection and processing of the respondents’ evaluation. Importance-Performance analysis (IPA) (Song et al., 2006), intuitive and operational but limited to the integrity of the destination image, is one of the main structural methods. Non-structural methods including but not exclusive to qualitative research and applying specific descriptions of interviewees in open questionnaires to obtain image information of tourism destinations, may work well in macroscopic researches on the overall image of tourism destinations as such methods are conducive to a deep and comprehensive destination image analysis; however, such analysis outcome may not be generalizable to different types of destinations. In conclusion, due to the complexity of tourism destination image perception and to overcome the mentioned shortcomings, we decide to combine structural and non-structural techniques in this specific study.

Whether the research method is structural or non-structural, researchers’ main data collection methods have conventionally been questionnaire surveys or interviews. These methods have merits in terms of verifying specific mechanisms and causal relationships but are faced with the dual challenge of sampling bias and timeframe constraint of conducting the surveys (Liu et al., 2019b). Improving internet access and increasing number of online travel websites coupled with the open, free, and sharing nature of the Internet Era have provided tourists with more convenient and efficient channels for information acquisition and dissemination (Zhang and Luo, 2017), thereby generating a massive new source of electronic data (UGC). UGC can include tourists’ reviews of specific scenic spots and tourist destinations. Due to their conciseness, timeliness, and volume, electronic reviews can reflect large number of tourists’ perceived image of tourism destination comprehensively. Zhang and Li (2014) compared two different data collection methods, questionnaire surveys and online reviews, by analyzing the differences between perceived and actual results of tourist destination image, and found relative uniformity in results between these two methods.

Present research on image destination based on UGC can be divided into two types. The first type is the overall destination image and characteristics analysis. Choi (2007) collected information related to tourism of Macau, China on Google and Yahoo’s website, and classified it into five main categories to analyze the tourism image of Macau. Fu et al. (2012) used ROST Content Mining software to extract featured words with high frequency from online reviews to analyze the content and uncover themes in tourist perception of Gulangyu, China. They also compared the image differences between different tourism destinations or the same tourism destination in different source markets. Stephenkova and Morrison (2006) discovered that the travel destination images of the same tourism destination in different source markets are different by conducting a comparative study of the travel agency websites in the United States and Russia. Some scholars also applied quantitative methods to explore the relationship between destination image and tourist-related factors. Zhang et al. (2014) applied a meta-analysis study, through which 14 hypotheses were developed and a total of 66 independent studies were synthesized and analyzed, to uncover correlations in the relationship between destination image and tourist loyalty. It is evident from these recent works that researchers are increasingly prone to combine various methods in their research in order to grasp a better understanding of the image of tourist destinations. For instance, a mixed-method approach including spatial analyses and cartography was employed to study overlap in the territorial distribution and attractiveness of tourism images distributed via official tourist brochures, travel guides, and user-generated content in Montevideo, Uruguay (Daniel, 2018).

Although related researches in this field are increasingly common at home and abroad in recent years, most of the researches are still applying only ROST Content Mining software to acquire the analysis of high-frequency words and represent the semantic network diagram, which is simple but not effective nor accurate enough. To date, current research has not yet developed a comprehensive technique to take full advantage of UGC to analyze the perceived image of tourism destinations. The present sentiment analysis based on travelers’ online reviews is mainly on the basis of a general lexicon that is not specific to tourism studies (Liu et al., 2019b). Therefore, there is little breakthrough in research methods and perspectives.

---

1 5A scenic spots: The scenic spots of People’s Republic of China are divided into five levels. The 5A scenic spot is the highest level of China’s tourist attractions, representing the world-class quality of China’s tourist attractions.
2.4. A critique

Firstly, from what had been mentioned above, it is clear that the comparative study of the projected image and the perceived image is of great significance in both the studies of tourist destination image and the practice of city promotion. The consistency of perceived images and projected ones is conducive to a memorable tourism experience for tourists as well as the development of destinations. Whereas, the mismatch between these two factors may exert a negative impact on the destination image, and cause serious damage via electronic ‘word-of-mouth’ images and the sustainable development of destinations. Therefore, it is necessary to delve into this topic.

Secondly, some combinations and modifications to analysis models and techniques have the potential to give rise to a more comprehensive understanding of the image gaps. Among all the relevant studies, the tourist destination image framework, the “cognitive-affective” model, proposed by Baloglu and McCleary (1999), is relatively mature and commonly in use, but it has not been applied in the comparative study of tourist destination image. In this paper, we employed this framework and integrated it with the analysis of projected and perceived destination image, to investigate the image gaps. We analyzed the cognitive images by referring to the division logic of Beerli and Martin (2004) from three dimensions: natural and humanistic resources, tourism reception and service facilities, and social environment atmosphere. Then, the affective image is based on the opposite emotion ring model proposed by Feldman and Russel (1998), and analyzed from the perspective of positive and negative emotions. Aiming at the official projected images, the current research mainly focuses on non-structural methods, and conducts qualitative analyses of official promotional materials, including traditional paper-made materials and updated channels such as electronic materials (Liu, 2015). The techniques of measuring perceived image include structural and non-structural methods, both of which have advantages and disadvantages. Existing researches tend to focus on one of them, but we believe combining both methods will yield better outcomes.

With the rapid proliferation of electronic platforms and the gradual consumer shift to digital lifestyles, the major information channels available to tourism studies have changed tremendously in recent years. UGC is increasingly becoming mainstream in contemporary tourism destination image studies. Despite this, the existing research has not comprehensively exploited the benefits of tourism UGC yet. We advocate that the methods in analyzing UGC need to be improved given the huge potential in the large amount of subjective and qualitative UGC in tourism studies still underexplored. Aiming to uncover characteristics of tourism activities and destinations, Liu et al. (2017a, 2019b) created a tourist destination image evaluation model based on the natural language processing method and Tourist Sentiment Evaluation (TSE) model. This is a tourist destination image measurement model that effectively combines qualitative and quantitative methods. It has been verified and applied to the study of contrasting Chinese tourists’ affective image with Australian tourists and to combine reputation and IPA to comprehensively evaluate the overall tourist perception of Guangdong destination. (Liu et al., 2017b, 2019a). We will be analyzing the tourist perception model based on the TSE model. Consequently, an analysis framework that integrates the “cognitive-affective” model of Baloglu and McCleary (1999) with the gap analysis of projected and perceived destination images, and the TSE model of Liu et al. (2017a, 2019b) will be employed to measure destination images.

![Fig. 1. Analysis logic and framework.](image-url)
3. Methodology

Our case study is Sanya, which is the only tropical island in China located on the southern side of Hainan Island. As of 2019, Sanya had received $2.4 \times 10^7$ visitors throughout the year and generated more than $6.0 \times 10^7$ CNY in tourism revenue. It has a well-defined destination image and positioning as “a casual, romantic, and relaxing resort city”, a perception that sits well among tourists. In view of this, Sanya is suitable for this study with as it has a developed tourism industry and perceivable tourism destination image.

Relying on the “cognitive-affective” model of Baloglu and McCleary (1999), the destination image, both for projected image and perceived image, was divided into two parts: cognitive component and affective component. In the cognitive component, based on the work of Beerli and Martin (2004), the analysis contained three dimensions: natural/cultural resources, tourist leisure infrastructures, and social environment setting. And the affective component was divided into two parts: positive and negative sentiment. This article will compare the two parts of the projected and perceived images one by one. The analysis logic and framework of this study is shown in Fig. 1.

The official projected image research is based on the analysis of the tourist attractions’ promotional videos from 2011 to 2016 (with the exception of 2014 video as it could not be found). The video’s visual and text content was analyzed to discover the cognitive and affective keywords for the official destination image analysis.

We conducted the analysis of tourists’ perceived image based on short reviews by tourists on Chinese online travel websites. The data collection included tourism reviews from five popular inbound travel websites, including http://lvyou.baidu.com, www.mafengwo.cn, www.tripadvisor.cn, www.qunar.com, and www.ctrip.com, to ensure the data is comprehensive and representative. Using a web crawler software, we collected 139,276 comments with about 4,500,000 words from reviews spanning from January 2011 to September 2016. We applied the high-frequency words to analyze the cognitive component of images and employed the TSE model to analyze the affective component, thereby using both above-mentioned methods in our study. ROST Content Mining software was first applied to extract the high-frequency words from collected online comments of Sanya. In order to avoid the loss of tourism lexicons, we added two phrases, namely tourism attraction and site name, after a plethora of tourism comments and over 50 individual travel notes of Sanya being read before the formal analysis (see Table 2). In addition, high-frequency words that were unrelated to our research purpose was omitted.

In terms of tourist perception, this study used the TSE model to calculate the affective aspects of reviews. Our method consisted of a word-frequency statistical method based on semantic logic and emotional preference correction, which largely includes these three parts: segmentation of reviews based on the tourist sentiment, exclusive use of thesaurus in the model; setting semantic logic rules to calculate the sentiment tendency of the review; and correcting the positive tendency of the review by the selected emotional multiplier (Liu et al., 2017a). Thereafter, a co-occurrence analysis was conducted respectively among high-frequency words of positive and negative reviews. This study also employed a clustering analysis program and visualization tool called Gephi to generate cluster networks, comprising a series of interconnected nodes, which represented the most-mentioned keywords among tourists’ review of Sanya. The size of the node denotes the importance of the words. The line connecting the nodes represents the association between keywords, whereas the distance between the nodes demonstrates closeness.

4. Results

4.1. Projected image of Sanya

4.1.1. Cognitive component of Sanya’s projected image

Table 3 demonstrates that in the cognitive component of official publicity materials, materials mainly present tropical and natural resources, sports activities, fashion shows, famous events, national culture, health resort, and high-end amenities. These are the tourist attractions that Sanya DMOs highly recommend with the aim of familiarizing tourists with them.

4.1.2. Affective component of Sanya’s projected image

The DMOs tend to convey only positive emotions with no negative emotions in promotional videos and Sanya is no exception. Specifically, Sanya’s official publicity team makes efforts to convey such a specific affective image to entice tourists to Sanya by portraying it as a beautiful, trendy, energetic, romantic, modern, relaxing, healthy, and cultural resort town. As can be seen from Table 4, “healthy” and “relaxation” have always been the focus of publicity by Sanya’s official agencies, and these image portrayals have remained relatively stable for some time. However, there is a shift in the portrayal of Sanya’s image since 2013 and official agencies started promoting “romance” instead. The position shift is reflected in the type of major events held in Sanya. Sanya has been the leader in hosting international event in China since 2003. In 2005, different types of classic events were hosted in Sanya at the same time. But at the end of 2010, after a release of the popular domestic romantic movie, Sanya has gradually become synonymous with romance and since then, has been working closely with the wedding industry.

\[\text{2011:}\ \text{https://v.youku.com/v_show/id_XMzI2NjMwNzA0.html?spm=a2hbt.13141534.0.13141534.}\ \text{2012:}\ \text{https://v.youku.com/v_show/id_XNDc0NTY0NDc2.html.}\ \text{2013:}\ \text{https://v.youku.com/v_show/id_XNjI3OTczNDU2.html.}\ \text{2015:}\ \text{https://v.youku.com/v_show/id_XMTQwMDc5ODA5Mg==.html?spm=a2h0c.8166622.PhoneSokuUgc_3.dtitle.}\ \text{2016:}\ \text{https://v.youku.com/v_show/id_XMTU5NTIzNDc4OA==.html?spm=a2h0c.8166622.PhoneSokuUgc_3.dtitle.}\]
4.2. Perceived image of Sanya

After processing the data, the high-frequency keywords of cognitive and affective images in 139,276 online reviews were extracted and their occurrence frequency was calculated, giving rise to the semantic network graph of positive and negative comments being constructed.

Table 2
Additional words for tourism-specific lexicon.

<table>
<thead>
<tr>
<th>Type</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tourism attraction</td>
<td>Sunshine, hibiscus, butterfly, sea, beach, Maldives, mangrove forest, the Li ethnic groups, Buddhism, Yazhou, love, honeymoon, “If You Are the One 2”, diving, motor boat, photograph, seafood, Wenchang chicken, Baoluo rice-flour, coconut</td>
</tr>
<tr>
<td>Place name and scenic name</td>
<td>Great East Sea, Nanshan, The Remotest Corners of the Globe, Yalong Bay, Wuzhizhou Island, duty-free shop, Guojianglong Bridge</td>
</tr>
</tbody>
</table>

Table 3
The keywords for cognitive component in Sanya’s tourism promotional video.

<table>
<thead>
<tr>
<th>Year</th>
<th>Keywords for cognitive component</th>
<th>Natural and cultural resources</th>
<th>Tourist leisure infrastructures</th>
<th>Social environment setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>Sea, sunshine, beach, coconut tree, temple, bridge, forest, Guanyin</td>
<td>Yacht, sailboat, wakeboard, luxurious dining, massage, yoga, national handicraft, events, duty-free shop, boat, golf, beach volleyball game</td>
<td>Shopping, photo taking, family, holiday resort</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>Sunshine, beach, sea, coconut tree, climate, rainforest, air quality, coral reef, bay, tropical island, ethnic customs, scenic spots</td>
<td>Airline, tourist real estate, shopping, sports games, beauty contests, events and festivals, international brand hotel, national dining, golf, yacht, SPA, high-end conferences</td>
<td>Ethnic minorities, family, holiday resort, foreign people, mega-events, marine sports</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>Scenic spots, tropical climate (sunshine, beach, sea, coconut tree, rainforest), island scenery, folk-custom, delicacy</td>
<td>Cross-country adventure, motorboat, teaching, golf, events, international brand hotel, sports</td>
<td>Couple, wedding, family activity, holiday resort, mega-events, Marine sports</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>Tropical climate (sunshine, beach, sea, coconut tree, rainforest), flower, air, bays, ethnic customs, scenic spots, seafood, tropical fruits</td>
<td>International brand hotel, international cuisine, golf, sports, duty-free shop, SPA, yacht, sports games, beauty contests, events and festivals, high-end conferences</td>
<td>Tropical paradises, holiday resort, family, mega-events, foreign people</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>Forest, sea, animals, flower, traditional culture, Nanshan Guanyin, delicacy</td>
<td>Sports, underwater restaurant, Qianguqing, seafood, yoga, show, events and festivals, golf, sports, duty-free shop, SPA, yacht</td>
<td>Couple, wedding, foreign people, mega-events, tropical paradises, holiday resort</td>
<td></td>
</tr>
</tbody>
</table>

Table 4
Keywords for affective components in Sanya’s tourism promotion videos.

<table>
<thead>
<tr>
<th>Year*</th>
<th>Keywords for affective component</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>Relaxation, breeze, freshness, tranquility, breath, dream, time, heart, satisfaction, happiness, smile, femininity</td>
</tr>
<tr>
<td>2012</td>
<td>National, tropical, beautiful, mysterious, unique, culture, quality, tasty, romantic, fashion, healthy, longevity, livable, vitality, openness, relaxation, vacation</td>
</tr>
<tr>
<td>2013</td>
<td>Tropical, charming, romantic, scientific, vacation, luxury, high-end, vitality, relaxation</td>
</tr>
<tr>
<td>2015</td>
<td>Tropical, beautiful, full of culture, international, relaxation, openness, fashion, vacation, romantic, high-end, vitality, hospitable</td>
</tr>
<tr>
<td>2016</td>
<td>Relaxation, ease, nature, passion, tradition, taste, love, femininity, fashion, happiness, romantic, international</td>
</tr>
</tbody>
</table>

Note: *, with the exception of 2014 video as it could not be found.

4.2. Perceived image of Sanya

After processing the data, the high-frequency keywords of cognitive and affective images in 139,276 online reviews were extracted and their occurrence frequency was calculated, giving rise to the semantic network graph of positive and negative comments being constructed.

Table 5
High-frequency word and frequency of natural and cultural resources.

<table>
<thead>
<tr>
<th>No.</th>
<th>High-frequency word</th>
<th>Frequency</th>
<th>No.</th>
<th>High-frequency word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Beach</td>
<td>21,858</td>
<td>12</td>
<td>Sunshine</td>
<td>2061</td>
</tr>
<tr>
<td>2</td>
<td>Sea</td>
<td>20,904</td>
<td>13</td>
<td>Reef</td>
<td>1470</td>
</tr>
<tr>
<td>3</td>
<td>Landscape</td>
<td>12,849</td>
<td>14</td>
<td>Blue sea</td>
<td>1260</td>
</tr>
<tr>
<td>4</td>
<td>Guanyin</td>
<td>10,929</td>
<td>15</td>
<td>Maldives</td>
<td>1179</td>
</tr>
<tr>
<td>5</td>
<td>Coral</td>
<td>4044</td>
<td>16</td>
<td>Buddhism</td>
<td>939</td>
</tr>
<tr>
<td>6</td>
<td>Stone</td>
<td>3963</td>
<td>17</td>
<td>Visibility</td>
<td>687</td>
</tr>
<tr>
<td>7</td>
<td>Roundabout</td>
<td>3567</td>
<td>18</td>
<td>Tropical paradise</td>
<td>582</td>
</tr>
<tr>
<td>8</td>
<td>Park</td>
<td>3444</td>
<td>19</td>
<td>Underwater world</td>
<td>492</td>
</tr>
<tr>
<td>9</td>
<td>Seacape</td>
<td>3093</td>
<td>20</td>
<td>Coastal</td>
<td>474</td>
</tr>
<tr>
<td>10</td>
<td>Blue sky</td>
<td>3018</td>
<td>21</td>
<td>Movie</td>
<td>408</td>
</tr>
<tr>
<td>11</td>
<td>Water quality</td>
<td>2145</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2.1. Cognitive component of Sanya’s tourists perceived image

The high-frequency words of cognitive image can be divided into three parts: natural/cultural resources; tourist leisure infrastructures; and social environment setting. Only the high-frequency nouns and pronouns which appear over 100 times and related to Sanya’s cognitive image were selected.

As shown in Table 5, the reviews related to natural/cultural resources are mainly classified into three categories: natural resources, environmental climate, and cultural resources. The words depicting natural resources, like “seawater”, “beach”, “blue sea”, “coral”, and “coast”, demonstrated that tourists have a high familiarity with Sanya’s unique tropical coastal resources. The frequency of “Maldives” is up to 1179, which confirms the superiority of Sanya’s natural resources in tourists’ minds, and also conforms to Sanya’s nickname—“Maldives of China”. The words that describe environmental climate, such as “sunshine”, “blue sky”, and “tropical paradise”, are mentioned about 3000 times in the reviews. This shows that tourists were able to experience the tropical charm of Sanya. In terms of man-made attractions, there are “Guanyin”, “Buddhism”, and “movies”, among which the frequency of “Guanyin” is as high as 10,929 times, revealing that tourists’ perception of Nanshan Guanyin in Nanshan Tourist Culture Zone is very high. It is noteworthy that the high-frequency words related to cultural resources do not include the Li ethnic groups culture, which is highlighted in the official promotion videos, indicating that the portrayed Li ethnic groups culture has not left a deep impression on tourists.

It can be seen from Table 6 that the high-frequency words related to hospitality and service infrastructures may be divided into three sub-groups of infrastructures: accommodation, transportation, and entertainment. Firstly, the high-frequency words of accommodation facilities include “hotel” and “five-star”. The former is mentioned 5777 times, which indicates to a certain extent that the most popular accommodation choice for tourists in Sanya is hotel accommodation. Furthermore, “five-star” indicates that generally speaking, hotels in Sanya are perceived to be high-end and luxurious. Secondly, the characteristic words of entertainment facilities include “diving”, “speed boat”, and “motorboat”, and those of recreational activities are “sightseeing”, “touring”, “swimming”, etc, which suggests that there are a great variety of marine entertainment facilities and recreational activities in Sanya. Finally, words concerning transportation facilities include “dock”, “boat riding”, and “bus”. However, the reviews barely mentioned food and beverage (F&B) facilities, with only food-related terms such as “seafood” and “package” appearing. It is apparent from the reviews that to date the tourists’ main F&B choices in Sanya are mainly cooked seafood, home-style dishes and the high-end F&B outlets of five-star hotels. There are few restaurants with local characteristics and of a certain scale, possibly indicating that there is an urgent need to upscale or upgrade the F&B industry in Sanya.

As shown in Table 7, the high-frequency words related to the social environment setting in Sanya may be divided into three main categories: city atmosphere, ethnic or ‘local’, and overall tourist perspective. It can be seen from the words, “resort” and “photographing”, that Sanya is a relaxing seaside resort in the eyes of visiting tourists. The high-frequency words, “couples” and “lovers” illustrate the romanticism of Sanya, the leisure tourism city. However, existing studies believe that “romanticism” is only an official gimmick by the Sanya government (Feng, 2016), which has not been confirmed.

Nevertheless, negative words, like “cheats”, “commercialization”, and “commercial atmosphere”, uncover the underlying problems of Sanya’s business environment during tourism development, which have left a deep but detrimental impression in the eyes of the tourists.

4.2.2. Affective component of Sanya’s tourists perceived image

After understanding tourists’ cognition of Sanya’s image, this study further analyzed Sanya’s affective component of destination images. We obtained 100,778 positive and 30,908 negative reviews by rating the positive and negative reviews, respectively. The ratio of positive vs. negative reviews is approximately 10:3. Thus, it is apparent that tourists generally have strong and positive emotions towards Sanya.

As shown in Fig. 2, there are 45 nodes and 98 edges in the positive reviews semantic network diagram. The positive reviews can be segregated into four subject clusters displayed in four different colors. The first cluster (in pink) is centered around “Sanya” and “site”, and mostly includes Sanya’s beautiful seaside scenery and tourism amenities such as tickets and transportation, which is the integral evaluation of tourists. The second cluster (in green) is centered on the “sands” and “seawater”, and mainly involves tourists’ specific positive feelings towards Sanya’s core attractions, including clear and clean water, fine sand, comfortable sunshine at the seaside, diving projects, hotels at the seaside, etc. The remaining two clusters (in blue and yellow, respectively) are relatively small and located at the edge of the network diagram, centered on “Yalong Bay” and “Guanyin”, respectively, representing two well-known scenic spots in Sanya.

Based on the size of point, the positive reviews semantic network consists of five core words: “Sanya”, “beach”, “seawater”, “local area”, and “Yalong Bay”. The central keywords and densities can be interpreted as below:

Table 6

<table>
<thead>
<tr>
<th>No.</th>
<th>High-frequency word</th>
<th>Frequency</th>
<th>No.</th>
<th>High-frequency word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hotel</td>
<td>5777</td>
<td>9</td>
<td>Bus</td>
<td>510</td>
</tr>
<tr>
<td>2</td>
<td>On the island</td>
<td>6281</td>
<td>10</td>
<td>Speed boat</td>
<td>346</td>
</tr>
<tr>
<td>3</td>
<td>Diving</td>
<td>4563</td>
<td>11</td>
<td>Powerboat</td>
<td>321</td>
</tr>
<tr>
<td>4</td>
<td>Seafood</td>
<td>1487</td>
<td>12</td>
<td>Accommodation</td>
<td>265</td>
</tr>
<tr>
<td>5</td>
<td>Sightseeing</td>
<td>1142</td>
<td>13</td>
<td>Package</td>
<td>251</td>
</tr>
<tr>
<td>6</td>
<td>Tour</td>
<td>930</td>
<td>14</td>
<td>Five-star</td>
<td>193</td>
</tr>
<tr>
<td>7</td>
<td>Boat ride</td>
<td>828</td>
<td>15</td>
<td>Swimming</td>
<td>191</td>
</tr>
<tr>
<td>8</td>
<td>Dock</td>
<td>722</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7
High-frequency words and frequency of social environment setting.

<table>
<thead>
<tr>
<th>No.</th>
<th>High-frequency word</th>
<th>Frequency</th>
<th>No.</th>
<th>High-frequency word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Resort</td>
<td>4407</td>
<td>5</td>
<td>Commercialization</td>
<td>1313</td>
</tr>
<tr>
<td>2</td>
<td>Photography</td>
<td>4257</td>
<td>6</td>
<td>Commercial atmosphere</td>
<td>1210</td>
</tr>
<tr>
<td>3</td>
<td>Couples</td>
<td>1878</td>
<td>7</td>
<td>Overcrowding</td>
<td>500</td>
</tr>
<tr>
<td>4</td>
<td>Lovers</td>
<td>1653</td>
<td>8</td>
<td>Cheats</td>
<td>74</td>
</tr>
</tbody>
</table>

Fig. 2. Semantic network of positive reviews.

Fig. 3. Semantic network of negative reviews.
Firstly, the two keywords of natural resource and attractions appear most frequently in the reviews, from which it can be seen that tourists' positive feelings of Sanya primarily originate from its natural resources.

Secondly, “Yalong Bay”, “Nanshan Temple”, “The Remotest Corners of the Globe”, “Nanshan Guanyin”, “Great East Sea” and other tourist attractions are the main contributors to tourists’ positive perceptions. It can be seen that natural resources play a significant part in Sanya’s tourist attractions, and its beautiful seascapes, spectacular cultural, religious landmarks, and natural forest parks exert a positive impact on tourists’ emotions.

Thirdly, natural resources such as “beach” and “seawater” give visitors the strongest sense of positivity. The beaches and seawater in Sanya are of high quality compared to other similar places in other domestic destinations. Positive adjectives of “sands” and “seawater”, “suitable” and “beautiful” illustrate that tourists have positive feelings about the quality of the beaches and the beauty of the sea in Sanya.

Finally, “convenience” and “traffic” are at the outermost periphery of the network demonstrating that they are marginal factors. This indicates that they are not the main determinants of tourists’ positivity towards Sanya.

From Fig. 3 it can be seen that, there are 37 nodes and 89 edges illustrated in the negative reviews semantic network diagram. The negative reviews can be divided into three subject clusters displayed in three different colors. The distribution of these three clusters is centered around Sanya.

The high image recognition: for one thing, Sanya’s romantic urban atmosphere to the city; for another, in the aspect of a positive semantic network, it is apparent that the romantic love

4.3. Comparison between tourist perceived image and of official projected image

Compared with the official publicity, it could be deduced that tourists’ perception of Sanya is basically in accordance with the official agencies’ positioning of “Beautiful Sanya, Romantic Tianya” in terms of the overall image. Tourists’ positive perception of natural resources, like Sanya’s beautiful tropical coastal scenery, is the strongest. Additionally, it is highly perceived that Sanya is a paradise for holidaymakers and lovers, with words such as “lover” and “lucky” directly presenting a romantic city image. However, Feng (2016) dismissed “romantic” as an official gimmick by local government agencies but lacking in public recognition. Two factors may account for the high image recognition: for one thing, Sanya’s pleasant seaside environment and suitable base for wedding photography bestow a romantic urban atmosphere to the city; for another, in the aspect of a positive semantic network, it is apparent that the romantic love story-themed movie shot in Sanya is also a significant contributor to its perceived romantic image. During the film screening period, in order to take the advantage of the popularity of the film, the Sanya provincial government spared no effort to evoke tourists’ emotional resonance with romantic movie plots and love stories, with the official website even setting a “wedding honeymoon” section to promote these activities.

Moreover, high-end hotel facilities constitute an important part of positive perception image, instead of negative ones as inferred from the frequent reports. What must be highlighted is that “priceless seafood” and other F&B-related negative words, widely reported in the media, are not the core of the negative perception. Two possible explanations might account for this finding. Tourists might have prepared themselves mentally for these unpleasantnesses before sightseeing, thereby mitigating the rise of negative emotions. For another, strict government regulations and corrective actions might have improved the situation; however, the media is still reporting on past negative events.
Nevertheless, we still detected a mismatch between tourist perception and official publicity in destination images, mainly regarding tourists’ attitude towards core scenic spots and the significance of festival activities. Although the core scenic spots have been vigorously publicized on the official website, and downtown Sanya is highly recognized by tourists, tourists maintain a love-hate relationship with the popular tourist attractions. The main reasons for this ambiguity include tourists paying high admission prices for entry into popular scenic spots in Sanya but upon visiting these scenic spots, these tourists often find that the attractions are unworthy of the high ticket-prices paid and in the process, negative emotions regarding the ‘lack of value’ for the tickets arise. In addition, various events were held in Sanya in recent years. The fact that these high-profile events were rarely mentioned in the tourism reviews despite strong official publicity might be attributed to the short publicity period of these non-national events and their limited audience reach.

5. Discussion

This study aims to deepen the understanding of destination image by combining the information from the supply side (DMOs) and the demand side (tourists), optimizing the use of tourism UGC, and the new sentiment evaluation model of tourist destination image. Not only is this study an innovative attempt in identifying destination images through big data approach, but it also proposes a more refined framework for future related studies using big data. To begin with, in terms of research materials, short comments on tourism overcome the conventional limitations of traditional questionnaire surveys and network travel notes, such as question-orientation incompleteness or small sample size (Liu et al., 2017b). As demonstrated, the big data approach is a more comprehensive and conducive method to analyze tourist perception of destination image and main factors that lead to tourists’ negative or positive feelings about destinations. Additionally, this study breaks the limitations of existing researches that focus unilaterally on “cognitive-affective” destination images or on “projected-perceived” ones. By combining the “cognition-emotion” model with the analysis of the gap between projected and perceived destination images, and refining the cognitive image and affective image of Baloglu and McCleary (1999), more detailed insight into the contributors of the “projected-perceived” image gap is obtained and this benefits tourism operators to make better and more targeted decisions regarding destination management.

However, the findings in this study are subject to a few limitations. For example, the online reviews used in this study are from five mainstream travel websites and do not include internet-wide coverage. Future research is needed to cover more websites and further explore the differences between research findings stemming from online tourism reviews that are on different travel websites or different audience groups.

We believe that this study has four implications for improving the management of tourism destinations. Firstly, natural resources play a significant part in the semantic network of both positive and negative comments, indicating the great attention that tourists attach to them. It is imperative for Sanya to develop tourism and other industries in a sustainable way with the goal to maintain its outstanding natural resources to the greatest extent. Secondly, the positive effects of electronic word-of-mouth communication are supposed to create awareness to fellow tourists, and the kinds of tourism experiences that will bring positive or negative experiences to tourists through tourism reviews ought to be understood to correct image promotion methods and contents with the purpose of providing better experiences for tourists. Thirdly, we would like to suggest that tourism operators could look into developing some cultural tourism projects by exploring and utilizing destinations’ local culture, such as Sanya’s ‘Luobidong culture’, ‘Li ethnic groups culture’, and ‘ocean culture’. Finally, this study suggests Sanya to improve the management of the public environment and facilities, recruit more staff in popular scenic spots and establish a travel public platform to improve services and solve the problems of traffic congestion and long waiting times.

6. Conclusions

In summary, this study refined the “cognitive-affective” model and used visual analysis method to obtain the official projected image of Sanya. ROST Content Mining software, Gephi software, and TSE model were applied for data-mining Sanya tourists’ short reviews to find out about the image of tourists’ perception. We found that the perceived destination image of Sanya is basically consistent with the official positioning of “Beautiful Sanya, Romantic Tianya”. Tourists successfully perceive the official image of Sanya as “suitable for vacation” and “very romantic”. Furthermore, it turns out that the prices of F&B and accommodation may not play as significant a role in tourists’ negative perceptions as portrayed by the reports, possibly due to tourists’ mental preparation before traveling to Sanya, official regulatory actions and exaggerated media reporting. Despite these issues, the study has demonstrated that Sanya’s coastal tourism resources possess strong competitiveness in the domestic tourism market. Nonetheless, there is still a gap between official publicity and tourist perception when it comes to tourist attitude towards core scenic spots, the significance of festival events, and so on.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Thanks to Development and Application of Big Data Sentiment Assessment Model for Tourism: Temporal and Spatial Differences, National Identity and Cultural Interaction, supported by The Ministry of education of Humanities and Social Science Project (19YJAZH060), and Study on Agglomeration Pattern, Development Quality And Spatial Optimization Of Urban Leisure Industry In Guangdong-Hong Kong-Macao Greater Bay Area, supported by Guangdong Philosophical and Social Sciences Project (GD20SQ21).