Sentiment Analytics of Chinese Social Media Posts

Jinyan Chen*
Griffith Institute for Tourism
Griffith University, Queensland
Jinyan.Chen@griffithuni.edu.au

Susanne Becken
Griffith Institute for Tourism
Griffith University, Queensland
s.becken@griffith.edu.au

Bela Stantic
Institute of Integrated and Intelligent
Systems
Griffith University, Queensland
B.Stantic@affiliation.org

ABSTRACT

The growing number of social media users and their posts provide valuable data about the sentiment that they have toward different services as well as people. Recent advances in big data analytics and natural language processing provided means to automatically calculate sentiment in text. Sentiment analysis is method which can be used to analyze social media content, it basically converts social media post text into quantitative data. While significant work was directed toward sentiment analytics of English text there is limited attention toward sentiment analytic of Chinese language. In this work we propose and test method to identify sentiment in Chinese social media posts and to test our method we rely on posts sent by visitors of Great Barrier Reef by users of most popular Chinese social media platform Sina Weibo. We elaborate process of capturing, managing and also we describe method for sentiment calculation, which provided details of sentiment toward the different GBR destinations and demonstrate that the proposed method is effective to obtain information and to monitor visitors' opinion.

KEYWORDS

Sentiment Analytics, Social media, Natural Language Processing

1 INTRODUCTION

The growing role of social media is attracting increasing research interest. Social media plays significant role in many aspects of life including retails, politics, tourism, especially in information search and decision-making behavior, but also in tourism promotion and customer interaction, and electronic word-of-mouth [37]. A large number of users actively engage with social media, for example, Twitter has 313 million monthly active users worldwide[27]. Furthermore, over 1.94 billion people monthly use Facebook [38] and 700 Million users use Instagram [17]. People use social media to post stories from their daily life, and they are particularly likely to

*Corresponding author

share impressions or emotions related to their travel experience. For the tourism industry, it is therefore important to examine social media and understand what visitors share and how they perceive destinations, attractions and products.

The Great Barrier Reef (GBR) is the world's largest coral reef system stretching over 2,600 kilometers along the coast of Queensland and attracts over two million visitors each year from all around the world [5], [14]. Significant proportion of GBR visitors are from China. While Chinese visitors are very active users of social media they typically use Chinese platforms rather than those commonly used by other English speaking visitors. The most popular Chinese Social Media platform is Sina Weibo which has over 300 million active users [12]. To take advantage of this huge number of users and media posts on Weibo we decided to capture those posts which specifically talk about the GBR.

China is Australia's fastest growing inbound tourism market and largest contributor to international visitor's spending in Australia. By 2022 the number of Chinese visitors is forecasted to increase from 685,000 in 2014 to 1.4 million and their real inbound tourism expenditure will grow from 4.4 billions to 8.2 billions [4]. Understanding the Chinese visitor experience in Australia is therefore important. China has become the number one tourism market in the world in both number of outbound tourists and aggregated tourism expenditure [23]. Therefore it is very important to know what is experience of Chinese visitors about GBR and what their perceptions are. However, it is difficult to examine because of language barriers. It was found that 77% of Chinese visitors could not speak English well or not at all [4].

Recent advances in Big data analytics and natural language processing provided means to understand visitor experience and their sentiment toward certain services or places [26]. Sentiment analysis is method which can be used for analyzing social media content, it basically converts social media post text into quantitative data, it also can extract information about special events and patterns. Sentiment analysis is a challenging and computationally demanding task not just because of vast volume of data but also due to common grammatical errors and misspelling, slang and abbreviations. Additionally social media post can contain emoticons that carry valuable information for sentiment and should not be discarded.

There are many sentiment analysis methods for English text presented in literature such as Valence Aware Dictionary for Sentiment Reasoning (VADER) approach which is purposely developed for sentiment analysis of short text found in social media posts [16]. VADER relies on dictionary but also has set of rules, which takes into consideration punctuation, emoticons, and many other heuristics. Dictionary contains items with associated sentiment intensities which are annotated by humans. Because dictionary only contains English words, as a result VADER can only provide sentiment for

English posts while content written in other language is not taken into consideration for sentiment calculation. While the sentiment analysis is matured for English language there are limited work for Chinese language and it is mostly related to creation of dictionaries.

In order to assess Chinese visitor's experience, as they mostly rely on Chinese social media, in this work we propose and test method to identify sentiment in Chinese social media and as test we use posts that mentioned Great Barrier Reef (GBR) in Chinese or English language. We relied on most popular Chinese social media platform Sina Weibo Sina [25]. We elaborate process of capturing, managing, creation of dictionaries and we also describe method for sentiment calculation which provided details of sentiment toward different GBR destinations.

Additionally we provided sample of additional information as a proof of concept, which can be extracted from social media posts meta data; such as where from Chinese people originate who have interest and comment on GBR related posts or changes in number of post over different months.

2 RELATED WORK

Social media post have been harnessed for different purposes including monitoring environmental changes [6], [7] and Sentiment Analysis in Tourism [19], [21], [36]. In most cases concept relied on sentiment derived from short text of social media posts as well as analytics of posts' meta data along other available online or scientific data. With regard to sentiment analysis of English language literature elaborates methods which can be allocated into three main groups: Machine learning, Dictionary based, and Hybrid approaches. Typical representative of Supervised machine learning approach requires annotation of data as well as pre-processing, extraction of feature followed by learning process and finally classification [2]. Different statistical based machine learning, such as Support Vector Machine (SVM) and Naive Bayes, are some of methods used for sentiment analysis mentioned in literature [13].

Typical representative of Dictionary-based methods (also one we extended and relied in our previous work) is Valence Aware Dictionary for Sentiment Reasoning (VADER), it is suited for sentiment analysis of short text [16]. The VADER combines a dictionary and a series of intensifiers, punctuation transformation, emoticons, and many other heuristics to compute sentiment polarity in social media posts text. For instance, a dictionary may contain words, such as excellent, better, good, bad, worse, terrible, with their respective polarities.

Clear advantage of dictionary based methods for sentiment analysis is that there is no need for annotation of text for training, which is obviously less costly considering that the annotation and training for supervised machine learning method needs to be undertaken for each new context. Another advantage is possibility to create domain specific dictionaries which ensure higher accuracy of calculated sentiment scores. But there is still need to create dictionary and annotate polarity of words, however, it needs to be done only once. This is obviously less costly considering that the annotation for supervised machine learning method needs to be undertaken for each new context.

In hybrid approaches both dictionary and machine learning are used to independently compute sentiment polarities and then individual results are combined to provide sentiment polarity [18]. Specific hybrid model which is using keywords and Naive Bayes algorithm has been recommended to calculate sentiment polarities of social media tweets [8].

Recent literature also addresses the issues with regard to trust and reputation measures in social network systems, especially in presence of thematic social groups [11].

3 METHODOLOGY

A sentiment analysis process is composed of several independent steps as it can be seen in Figure 1. It starts with data collections which in case of social media data most often relies on utilization of dedicated available API's. In cases when purpose built Application Programming Interface (API) is not available web crawling is needed. Crawling initially involves an identification of a data source, for example, a commercial website or a social media network. Dedicated web crawling code needs to be developed and used to collect data from these sources. Considering that huge amount of post have been generated by users there is need to filter and acquire only relevant posts, most often filtering is related to particular geographic area or to particular keywords. After initial cleaning, which discards data that do not contain full records, data are saved in database.

Considering that the most relevant data are in JavaScript Object Notation (JSON) format NoSQL databases are best suited for data storing. Once data have been stored in a database, it is possible, for example, to calculate the sentiment from content of the text within a set of data fields in particular posts. In addition to the actual text of the post additional fields, if publicly available, can also be captured that form the meta data. If ths meta data is available it is possible to analyse locations where from posts have been sent or where from users originate, based on place where they opened account.

3.1 Data Extraction Process

In order to collect only relevant posts we created conditions which needs to be satisfied so posts are collected and stored in local NoSQL database. The following procedure was implemented. We extracted only all posts that contained the keywords "Great Barrier Reef" both in Chinese and English language.

To access and capture Weibo posts we utilized public Sina API, data is in JSON format and we stored it in locally in a Mongo NoSQL database, which was chosen because it has been shown that the relational databases are struggling in handling large amount of unstructured data [26]. NoSQl database is able to store and efficiently access a diversity of unstructured data including text, emotions and media files. Extraction process is shown in Figure 1. If the posts were associated with an identifiable GPS location, we separated them into another collection so they can be shown on a map (Latitude, Longitude). We identified that about 15% of posts have associated geo locations. In addition, the analysis involves assessment of the 'emotional' tone of the text; which will be calculated with method proposed in this paper.

Our experiments were conducted on in-house Big Data cluster running Hadoop (2.6.0) and MongoDB (3.2.9).
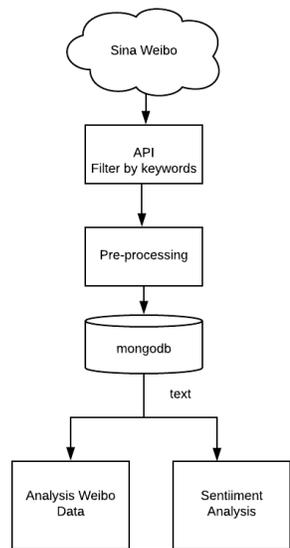
Figure 1: Global Sentiment analysis process of Weibo data.

3.2 Creation of Chinese Lexicon

Considering that there is limited work on sentiment for Chinese text, it is also reflected in shortage as well as quality of available Chinese language lexicons. Lexicon is dictionary of words along with their associated sentiment intensity measures, which are usually in the first instance validated by humans. For Chinese language there are limited number of dictionaries available.

When looking into available dictionaries we have identified that despite there are some duplication of words they complement each other. However biggest disadvantage was that they had no intensity of sentiment. Only lexicon from Bo Yuan, Tinghua University [1] has intensity associated with specific words however it had limited number of words. Other approach such as HowNet [40] had intensity associated with different dictionaries such as 'most', 'more', etc. Also dictionaries from Dalian University has a simple label for positive and negative polarity. All other dictionaries that we found are formed and grouped by positive or negative words.

In forming our lexicon we considered the following dictionaries:

- Chinese Sentiment Word Weight Table from BoYuan in Tinghua University: It contains 23,419 phrases with a positive or negative weight [1].
- How Net: has 17,887 phrases which are divided into 6 groups based on the phrases emotional tendency, which are "Positive Evaluation", "Negative Evaluation", "Positive Emotion", "Negative Emotion", "perception" and "Adverb of degree" [40].
- Chinese emotional vocabulary from Dalian University of Technology Information Research Laboratory. It contains 22,012 phrases [33].
- National Taiwan University Sentiment Dictionary - NTUSD: has both Simplified and Traditional Chinese Character. It has 2,810 positive words and 8,276 Negative words [30], [31].

- Positive and Negative dictionaries: contains 730 positive words and 939 negative words [39].
- Positive dictionary which has 5,067 positive words [24].
- Negative dictionary which has 3,495 negative words [34].
- Positive and Negative Dictionary which contains 4,468 negative words and 5,567 positive words [20].

Apart of being limited in content and number of words, as mentioned, these dictionaries have shortcoming as there is no sentiment intensity measures of individual words. Weight has been mostly addressed by creating dedicated dictionaries for specific words which are grouped based on association with: most, very, more, half, etc. Apart of lower accuracy of sentiment score this method would require dedicated code in programming for calculation of sentiment, which also would slow down the sentiment calculation.

Therefore we decided to create our own lexicon which will combine existing dictionaries and add weight intensity in accordance with meaning. For example we added weight +3 for words associated with *most* while weight +2 for *more*, etc. Also, we assigned weight +1 to words in positive dictionaries add -1 for words in negative dictionary. Words which existed in lexicons and had sentiment intensity validated by humans we retained such as from HowNet lexicon [10], [40]. As a result we created comprehensive lexicon for Chinese language with over 30,00 terms with associated weighting. We realize that the certain words still have only weight +1 or -1, however, it is at this stage sufficient and more accurate than any existing method and any existing lexicon. We plan to engage people to do the annotation of words sentiment weight, however, considering the fact that there are over 30,000 words and every individual could have different sentiment weight it is required to survey several people for same words, therefore this is lengthy and expensive task.

3.3 Sentiment Analysis of Chinese Short Text

Sentiment analysis, often referred to as opinion mining, is process that extracts the relevant information of public attitudes from the subjective text [35]. The purpose of this emotional analysis of the text is to identify polarity of the social media post, in order to gain insight whether posters have had positive, negative or neutral experience [2].

As indicated in Section 2. there are many sentiment analysis methods proposed for English language and they are mostly lexicon based. With regard to sentiment analysis of short Chinese language text there is limited work and process itself is more complicated when compared to sentiment analysis of English language, because text requires segmentation into meaningful words. In Chinese language there are no spaces between characters and specific meaning is often defined by combination of characters or words. Encouraged by work of our Big data group on sentiment analysis for English language we decided to follow same direction and propose lexicon based sentiment analysis method for Chinese language. In Figure 2 we show concept of sentiment analysis of Chinese text that has been proposed and developed in this work.

Calculation of sentiment was done according to the Algorithm 1. Once when Chinese social media posts *WEIBO* have been collected and stored in Mongodb databases in JSON format, with regard to sentiment calculation of posts first step is to only consider the
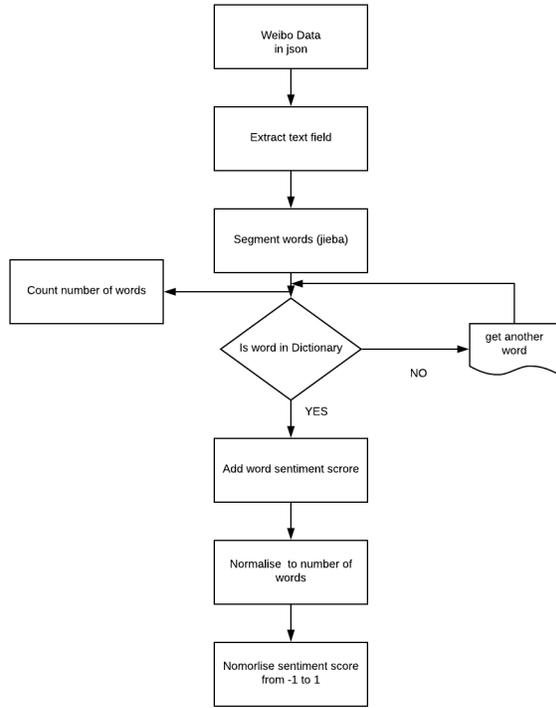
Figure 2: Sentiment analysis concept.

content of the posted text within the post *TEXT* and encode text with UTF-8 encoding *D*, suitable for Chinese language. Individual encoded text *d* is then considered for sentiment calculation. In the next step, considering that there is no space to separate characters and words in Chinese language, there is need to do the segmentation to meaningful words. In Chinese language, one character usually does not have meaning, which pose a challenge itself to segment text written in Chinese. For example, in English "I love traveling", "traveling" is one word , however, in Chinese language, it needs two characters "lv" and "you" to have the meaning of "traveling". Additionally, as there is no space to separate words in a sentence, if we need to find "traveling" in Chinese language it is required to take into consideration both "lv" and "you" in order to have the meaning of "traveling". Segmentation of Chinese language is research topic itself, and it is obviously outside of content of this work, it attracted significant attention in literature [15], [28], [3], [29], [22].

In this work on developing sentiment analysis of Chinese short text we accepted and followed *jieba* segmentation method [28]. Apart of being widely used *jeiba* has advantage because it has libraries for python, which we could simple plugged into python code for sentiment analysis. *Jieba* can search the maximum probability path and most probable combination based on the word frequency. It supports three word-segmentation models (accurate mode, full mode and search engine mode), can process the traditional Chinese word segmentation, and supports custom dictionaries [32]. *Jieba* segmentation libraries and all related files are publicly accessible at [28].

All segments from posts (*SEG*) are lopped and segments (*seg*) are matched with lexicon (*LEX*). If segment exist in lexicon associated intensity (*wordsentiment*) is obtained and added to the Sentence sentiment (*sen_sentiment*). Because post in Weibo can be short or very long, we count number of segments (*words*) as well as number of matching segments with lexicon (*dictwords*), which we take into consideration in calculation of sentiment of individual post. This is required because there are more positive than negative words in Lexicon and therefore it is more likely that the lengthy posts have more matching words with lexicon and therefore higher positive sentiment if adjustment was not performed.

Algorithm 1 Sentiment calculation of short Chinese text

WEIBO = Input Weibo posts in JSON format
LEX = Import Lexicon with sentiment intensity
$TEXT = Extract\_only\_text\_field\_from\_post(WEIBO)$
$D = encode\_with\_UTF8(TEXT)$
$sen\_sentiment = 0$
for all $d \in D$ **do**
    $words = 0$
    $dictwords = 0$
    $sentiment = 0$
    $SEG = Perform\_Segmentation\_of\_text\_with\_jeiba(d)$
    for all $seg \in SEG$ **do**
        $words = words + 1$
        if $seg \in LEX$ **then**
            $wordsentiment = LEX(seg)$
            $sen\_sentiment = sen\_sentiment + wordsentiment$
            $dictwords = dictwords + 1$
        end if
    end for
    $adjusted\_sent = sentiment * dictwords / words$
    $normalize\_to\_one = adjusted\_sent / \sqrt{adjusted\_sent^2 + 2}$

    Insert text and sentiment into database
end for

When all segments from individual post are taken into consideration adjusted at first adjusted sentiment (*adjusted_sent*) was calculated with the following equation:

$$adjusted\_sent = sentiment * dictwords / words$$

And finally post sentiment is normalized to 1 to ensure that the sentiment is always between -1 for negative and +1 for positive post. The following equation is used to normalize sentiment to 1:

$$normalize\_to\_one = adjusted\_sent / \sqrt{adjusted\_sent^2 + 2}$$

Normalized sentiment score is stored in database along the associated text and loop continued to next post until all posts are finished.

4 EXPERIMENTAL EVALUATION

In our experimental evaluation we considered Sina Weibo posted 2016 and 2017 that mentioned 'great barrier reef' both in English and Chinese language. A total of 24,308 relevant posts have been

Table 1: Sentiment Score of GBR related posts

Key words	SentimentScore2016	SentimentScore2017
Cairns	0.3828	0.4042
Townswille	0.385	0.4484
Green Island	0.4398	0.4406
Whitehaven Beach	0.3713	0.4423
Heart Reef	0.481	0.4747
Hotel	0.3261	0.4314
Seafood	0.4474	0.4043
Snorkeling	0.3551	0.3525

captured. Based on our previous research experience on great barrier reef [6], we selected few keywords including key locations (Townswille, Cairns, etc), food (Seafood), hotel and some relevant activities such as 'snorkeling' and than calculated overall sentiment for these keywords. As a demonstration of sentiment calculation in Table 1. we provide some sentiment scores calculated by method proposed in this work. It can be seen, for example that all related posts are positive in these 2 years, however, the overall sentiment about the travel destinations was increasing, while only seafood sentiment score dropped from 0.4474 to 0.4043.

Considering that posts apart of text social media posts can contain other relevant metadata we investigated what additional information metadata can provide. We noticed that out of 24,308 captured posts, almost all users (99%) provided their location at the time of registration, which could be good indication where the users are coming from. Figure 3. shows a heat map to illustrate where in China the Weibo users who talked about the GBR originate at the time of registration of their accounts. Based on this details we could identify number of users from different provinces, for example Beijng is top ranked (a total of 5,091 users) that mentioned Great barrier Reef, followed by Guangdong Province (2,389) and Shanghai (1,456). It is important to mention that this analysis result correlates well the results released by Queensland Tourism Industry Council which announced that "Markets in ChinaŠs first-tier cities (Beijing, Shanghai, Guangzhou and Shenzhen") [9]. It has been also previously demonstrated that the social media data correlates well with scientific observations [7]. Therefore, considering also observation in this work related to first-tier cities it is evident that the metadata can be used to gain a lot of valuable information.

It is interesting to note that there are posts from all provinces in China, indicating that the GBR is very popular in Chinese social media. Additionally, our sample contained posts from users who opened an account in Taiwan (150), HongKong (356) and Macao (81). Also there were 1,032 Weibo users who registered their account in Australia.

Figure 4. shows how many posts in different months in 2016. It is interesting to see that two peak times of posting about great barrier reef, which are February and September with increasing trend. In the Chinese visitor satisfaction report [9] says "Chinese holiday tourists arriving in Australia for leisure purpose have featured high seasonality. AustraliaŠs summer tends to be the peak season for Chinese tourists in order to avoid freezing winter. The peak season lasts for four months from November to February, covering several
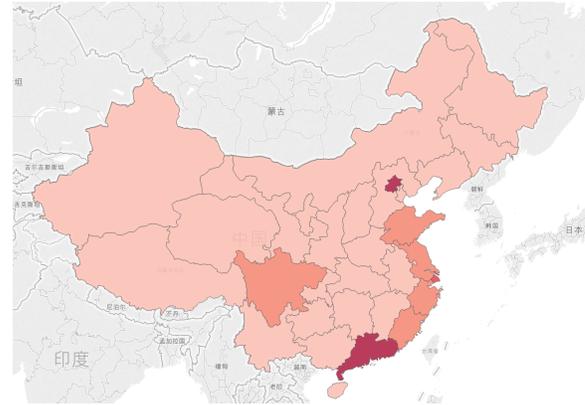
Figure 3: Heat map of user locations mentioned GBR from 2016 to 2017
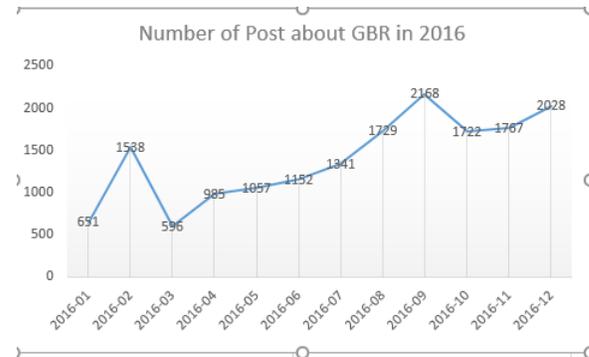
Figure 4: Number of posts per month in 2016

important holidays such as Christmas and New YearŠs Eve". Our finding correlates with this observation.

5 DISCUSSION

At the moment, Weibo has canceled the word limitation. It gives users chance to post more complicated and deeper ideas, which will make it more challenging to do the segmentation. Second, unlike Twitter which has different APIs (keywords and location) , Weibo API can use keywords to search, however at this stage does not support search by location. So a better program to access Weibo API will be required in the future. Also, to improve sentiment analysis result using lexicon based rules, a more comprehensive lexicon is needed, especially a lexicon that can include words that people like to use in social media, as well as emoji, need to be taken into consideration.

In the future, we will add more GBR related words, such as marine lives and activities in relation to the GBR. This will result in a larger number of posts, and also provide even further insight into who is talking about the GBR. Also, we will add an analysis of the location from where the post was sent, which can help to better understand visitors' behavior (including travel movements) and their specific needs. Mining weibo data has great potential in assisting the GBR-related tourism industry in understanding Chinese tourists and

their interest. Further it is possible to incorporate analysis to get insight into Chinese tourist profile with regard to the age, gender, and their interest as well as incorporate pictures and videos for analysis.

6 CONCLUSIONS

Knowledge of travel behavior and tourist sentiment is crucial to inform the tourism sector to plan for future growth, product development, and marketing. For the GBR visitor perceptions are also important in relation to the environmental management of the natural resource. Consumer perception assessments often rely on existing approaches to data collection such as surveys and opinion polls. However, these have a range of limitations both in terms of sample size and bias (i.e. accuracy). There is a risk that traditional 'paper and pen' methods are unable to fully capture opinions and behaviors. To overcome this limitation, we proposed to tap into available social media post and perform Big Data analytics. In order to gain insight into tourists with Chinese background we decided to capture weibo posts, the most popular Chinese social media platform. The findings provide insight into what locations are most talked about in weibo posts. Apart from the GBR in general, weibo users often referred to Cairns and Green Island - obviously two key locations that form part of the Chinese travel itinerary in the region. Analysing the origin of the users who discuss the GBR in their messages highlights where in China people are interested in the GBR. In addition to the provinces around Beijing and Guangdon, it was also found that a substantial number of weibo users registered their account in Australia. This finding deservers further research; this particular user group may reflect migrants or students who live in Australia but still maintain strong connections to China. Their word-of-mouth and opinions about the GBR could be influential in shaping an overall 'sentiment' about the Reef. It would be useful to find out more about the specific context and their followership, both within Australia and overseas. Sentiment analysis was applied to all text data; especially we compared the sentiment score for each city. The sentiment score reflects the opinion of visitors travelling experience at GBR, which can give advice for authority to estimate the tourists' satisfaction. Initial sentiment analysis reveals that the best sentiment is toward the experience on beaches, while sentiment is lower when talking about the experience on Islands, deeper analysis could reveal reasons for significant variation of sentiment.

We captured keywords relevant to tourism and Great Barrier Reef and harnessed Sina Weibo API's to obtain posts and do analysis of the sentiment as well as identify where from tourist are coming from and what are their interest. These finding can help tourism operators to attract more visitors as well as improve marketing strategies.

ACKNOWLEDGMENTS

REFERENCES

[1] [n. d.]. Chinese Sentiment Word Weight Table, BoYuan, Tinghua University. ([n. d.]). http://www.datatang.com/data/43216
[2] Ali Reza Alaei, Susanne Becken, and Bela Stantic. 2017. Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research* (2017), 0047287517747753.
[3] Ansjsun. [n. d.]. ([n. d.]). http://www.cnblogs.com/en-heng/p/6274881.html
[4] Tourism Research Australia. 28 January,2014. CHINESE VISITOR SATISFACTION. (28 January,2014). Retrieved 28 January, 2014 from https://www.tra.gov.au/Research/View-all-publications/All-Publications/International-market-reports/chinese-visitor-satisfaction
[5] Susanne Becken, Char-lee McLennan, and Brent Moyle. 2014. *World Heritage Area at Risk?: Resident and Stakeholder Perceptions of the Great Barrier Reef in Gladstone, Australia.* Griffith Institute for Tourism, Griffith University.
[6] S. Becken, B. Stantic, A. Alaei J. Chen, and R. Connolly. 2017. Monitoring the environment and human sentiment on the Great Barrier Reef: Assessing the potential of collective sensing.. In *Journal of Environmental Management*, Vol. 203. 87–97.
[7] Jinyan Chen, Sen Wang, and Bela Stantic. 2017. Connecting Social Media Data with Observed Hybrid Data for Environment Monitoring. In *Proceedings of the 11th International Symposium on Intelligent Distributed Computing - IDC.* 125–135.
[8] William B Claster, Malcolm Cooper, and Philip Sallis. 2010. Thailand–Tourism and conflict: Modeling sentiment from Twitter tweets using naïve Bayes and unsupervised artificial neural nets. In *Computational Intelligence, Modelling and Simulation (CIMSiM), 2010 Second International Conference on.* IEEE, 89–94.
[9] Queensland Tourism Industry Concil. [n. d.]. ([n. d.]). https://www.qtic.com.au/resources/chinese-visitor-experience-report-october-2013
[10] Professor Zhendong Dong. 2000. HowNet. (2000). http://www.keenage.com/html/e_index.html
[11] Chi Thang Duong, Quoc Viet Hung Nguyen, Sen Wang, and Bela Stantic. 2017. Provenance-Based Rumor Detection. In *28th Australasian Database Conference - ADC.* 125–137.
[12] Bo Fan. 2016. Weibo data report in 2016. (2016). Retrieved 2017 from http://data.weibo.com/report/reportDetail?id=346
[13] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford* 1, 12 (2009).
[14] Greatbarrierreef. [n. d.]. www.greatbarrierreef.org. ([n. d.]). Retrieved February 28, 2018 from http://www.greatbarrierreef.org/about-the-reef/great-barrier-reef-facts/
[15] Zhang Hua-ping and Liu Qun. 2013. ICTCLAS. (2013).
[16] C. Hutto and E. Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *In Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (2014).
[17] instagram. 26 April,2017. instagram. (26 April,2017). Retrieved 2017 from http://blog.instagram.com/post/160011713372/170426-700million
[18] Walter Kasper and Mihaela Vela. 2011. Sentiment analysis for hotel reviews. In *Computational linguistics-applications conference*, Vol. 231527. 45–52.
[19] Daniel Leung, Rob Law, Hubert Van Hoof, and Dimitrios Buhalis. 2013. Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing* 30, 1-2 (2013), 3–22.
[20] Jun Li. 22 January,2011. Chinese positive and negtive Lecxion. (22 January,2011). Retrieved 2011 from http://nlp.csai.tsinghua.edu.cn/site2/index.php/resources/13-v10%20Seeaustralia,%202017
[21] Zhan Liu, Jialu Shan, Nicole Glassey Balet, and Gang Fang. 2017. Semantic social media analysis of Chinese tourists in Switzerland. *Information Technology & Tourism* 17, 2 (2017), 183–202.
[22] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations.* 55–60.
[23] Tien Duc Pham, Son Nghiem, and Larry Dwyer. 2017. The determinants of Chinese visitors to Australia: A dynamic demand analysis. *Tourism Management* 63 (2017), 268–276.
[24] JiLin Shi and Guiying Zhu. 2005. *PositiveDictioanry.* SiChuanCiShu.
[25] Sina. 2009-2018. (2009-2018). https://weibo.com
[26] B. Stantic and J. Pokorný. 2014. Opportunities in big data management and processing. *Databases and Information Systems* 270 (2014), 15–26.
[27] statista. 28 January,2014. Number of monthly active Twitter users worldwide from 1st quarter 2010 to 4th quarter 2017 (in millions). (28 January,2014). https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/
[28] J Sun. 2012. ŚJiebaŠ Chinese word segmentation tool. *available at https://github.com/fxsjy/jieba* (2012).
[29] Maosong Sun, Xinxiong Chen, Kaixu Zhang, Zhipeng Guo, and Zhiyuan Liu. 2016. *Thulac: An efficient lexical analyzer for chinese.* Technical Report. Technical Report.
[30] National Taiwan University. [n. d.]. ([n. d.]). Retrieved 2018 from http://academiasinicanlplab.github.io/

[31] National Taiwan University. 22 January,2011. NTUSD: National Taiwan University Semantic Dictionary. (22 January,2011). Retrieved 20 June 2017 from https://rdrr.io/rforge/tmcn/man/NTUSD.html

[32] Kecong Xiao, Zishuai Zhang, and Jun Wu. 2016. Chinese text sentiment analysis based on improved Convolutional Neural Networks. In *Software Engineering and Service Science (ICSESS), 2016 7th IEEE International Conference on.* IEEE, 922–926.

[33] Linhong Xu, Hongfei Lin, Yu Pan, Hui Ren, and Jianmei Chen. 2008. Chinese emotional vocabulary. 27, 2 (2008), 180–185.

[34] Ling Yang and HuiYing Zhu. 2005. *NegativeDictioanry.* SiChuanCiShu.

[35] Fulian Yin, Beibei Zhang, Pei Su, and Juanfang Chai. 2016. Research on the text sentiment classification about the social hot events on Weibo. In *Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2016 IEEE.* IEEE, 1537–1541.

[36] F. Yin, B. Zhang, P. Su, and j. Chai. 2016. Research on the text sentiment classification about the social hot events on Weibo.. In *IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC).* 1537–1541.

[37] Benxiang Zeng and Rolf Gerritsen. 2014. What do we know about social media in tourism? A review. *Tourism Management Perspectives* 10 (2014), 27–36.

[38] zephoria. 2017. Valuable Facebook Statistics. (2017). Retrieved 2017 from hhttps://zephoria.com/top-15-valuable-facebook-statistics/

[39] Wei Zhang, Jin Liu, and XianZhen Guo. 2004. *StudentPositiveAndNegativeDictioanry.* China Encyclopedia Publishing House.

[40] Dong Zhendong and Dong Qiang. 2006. *Hownet And The Computation Of Meaning.* World Scientific.