Application of Data Mining in Sina Weibo — Sentiment Indicator to Gauge Tourist Satisfaction in Macao

Rita T. Tse

***Abstract*—Sentiment analysis and opinion mining are active research trends in data mining. The explosion of social media such as social networks has created unprecedented opportunities for data mining research community. Analyzers can study and analyze users' opinions, attitudes, and emotions about news or social events. Our focus in this work is to propose a way to gauge tourist satisfaction based on some criteria set forth in the ranking of China's Feature Tourist City, by reviewing Natural Resources, Tourist Goods, and History and Culture of the city from the social network. The tweets collected from Sina Weibo in Macao have gone through a series of process to reach the expected data format and then be eventually analyzed. A dataset of 418,056 tweets from 2013 and 194,880 tweets from 2014 is analyzed in this work. Tourist satisfaction is calculated and the result indicates that the tourists' overall satisfaction towards Macao is generally positive.**

***Index Terms*—Data mining, Macao, sentiment analysis, Sina Weibo.**

I. Introduction

With the rapid growth of World Wide Web and electronic commerce, huge volume of data is available online. The available online data is plenty and with rich descriptions. The Internet has provided a unique opportunity for business. About 2.9 billion of internet users were registered worldwide in 2014, almost half of them lived in Asia. About 48 percent of all internet users live in Asia, with China being the global leader with about 641 million of internet users [1]. Furthermore the Chinese urban population appears to be sensible to the adoption of new media such as social media as a mean to express the new level of lifestyle they achieved. The social network, such as Sina Weibo in China, has reached 222 million monthly active users as of September 2015 and trillions of data generated on the platform [2]. It becomes a huge platform that people can share their feeling to others.

Tourist satisfaction is an important qualitative indicator from the demand side. Customer demand is determined by a range of factors including the experience and recommendation of others. Social media, blogs and other review sources play an important role in the decision making of many travelers [3].

In this paper, we will discuss the application of data mining in the social network to gather opinion about how tourists feel about Macao, in terms of the three categories - Natural Resources, Tourist Goods and History and Culture of the city (some criteria used by the ranking in the China's Feature Tourist City [4] and Macao ranking no. 2 in 2014 [5]). The tweets from Macao in 2013 and 2014 collected will be used. Based on the sample of tweets, what are tourists' opinion about Macao?

Opinions might be positive, negative or neutral. This makes the mission for the analysis quite difficult. A Sentiment Analysis will be presented in this paper, which is the process of identifying positive and negative opinions, and emotions [6]-[8].

The outline of this paper is as follows: We start with a study of the related work on social network - Sina Weibo. Next we follow with the discussion of system architecture, data collection method and analysis methodology. Finally, we present the conclusion of the paper.

II. Related Work

Various types of research efforts have been conducted on Sina Weibo data ranging from user behavior [9]-[13] to other applications such as mining opinion [14], [15], detecting rumors [16], [17], and finding out travel experiences and preferences of places to visit [18], [19].

Sina Weibo, is designed as platforms allowing users to generate contents that open to the public. From analyzing posts submitted to Sina Weibo, some activities of users can be estimated. For example, Li *et al.* [9] found that leisure & mood and hot social events account for almost 65% of the popular topics discussed by Sina Weibo users and Yu *et al.* [10] demonstrates its possible role to detect the sleeping time of users and find a new method for judging users' time zone.

Research on user behavior on the Social Web has also been initiated. Chen and She [11] analyzed the Weibo social network with verifications, by comparing the user microblogging behaviors between verified users and unverified users and studying the social network evolvements of these two group of users. Empirical results showed that the verifications stimulated people to follow verified users and actively participate in the online social activities. Guan *et al.* [12] selected hot events that were widely discussed on Sina Weibo and found that male users are more likely to be involved and messages that contain pictures and those posted by verified users are more likely to be reposted. Furthermore Cui *et al.* [13] explored the relationships among different profile attributes in Sina Weibo by using association rule mining to identify the dependency among the attributes. The results found that if a user's posts are welcomed, he or she is more likely to have a large number of followers.

On one hand, a great number of people are making use of the social network to express their opinions, which also

Manuscript received January 17, 2016; revised March 22, 2016. This work was supported in part by the Macao Polytechnic Institute-Bridging Urban Sensing and Social Networks (RP/ESAP-02/2014).

Rita T. Tse is with the Computing Program, Macao Polytechnic Institute, Macao, China (e-mail: ritatse@ipm.edu.mo).

doi: 10.18178/ijimt.2016.7.2.649

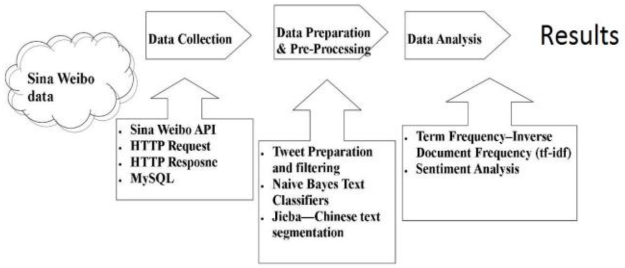
makes Sina Weibo an extremely valuable source for learning thoughts of crowds on given topics. One challenge is finding the important ones from a large amount of Weibo posts. Ren *et al*. [14] investigated approaches on selecting representative posts. Several methods including LexRank, Turn down the Noise (TDN) and Kmeans can be used to evaluate and select representative posts, in the research such as opinion summarization and information retrieval. In general, sentiment analysis is concerned with the analysis of direction-based text, for example, text containing emotions and opinion. Recently, Fan *et al.* [15] found that the correlation of anger among users is significantly higher than that of joy.

On the other hand, other type of research includes rumor detection in the social network. The automatic assessment of information credibility therefore becomes a critical problem, because there is often not enough resource to manually identify the misinformation about controversial and large scale spreading news from the huge volume of fast evolving data. Yang *et al.* [16] trained a classifier to automatically detect the rumors from a mixed set of true information and false information in Sina Weibo. Meanwhile, Wu *et al.* [17] proposed a graph-kernel based hybrid SVM classifier which captures the high-order propagation patterns such as topics and sentiments and achieved a classification accuracy of 91.3% on randomly selected Weibo dataset.

In addition, blogs and microblogs become popular in recent years because people like to share their travel experiences online. Tse and Zhang [18] analyzed blog and microblog contents created by mainland Chinese visitors sharing their Hong Kong experiences and research results indicate a generally positive image of Hong Kong as a destination among the mainland Chinese bloggers. Meanwhile Lao and Tse [19], studied tweets from Sina Weibo in Macao and Hong Kong, and found user preferences of various tourist places to visit.

III. Approach

This section presents the approach used for analyzing a huge amount of tweets from Sina Weibo in Macao to calculate tourist satisfaction, by reviewing Natural Resources, Tourist Goods and History and Culture of the city. The sub-sections below will explain the details of our approach which include: data collection, data preparation and pre-processing, data analysis and results, as illustrated in Fig. 1.



- Sina Weibo API
- HTTP Request
- HTTP Resposne
- MySQL

- Tweet Preparation and filtering
- Naive Bayes Text Classifiers
- Jieba—Chinese text segmentation

- Term Frequency–Inverse Document Frequency (tf-idf)
- Sentiment Analysis

Fig. 1. The system architecture for Sina Weibo tweets data mining.

A. Sina Weibo as Data Corpus

Because of the popularity and the huge amount of data created in Sina Weibo, it was selected as the main source and input for this work.

B. Data Collection

A data collection program is developed to collect tweets from Sina Weibo users in Macao every 5 minutes in 2013 and 2014. The steps are to 1) use the Sina Weibo API, 2) get the update access token and 3) send HTTP Request to Sina Weibo Server using the access token. To send the HTTP request to Sina Weibo, the GPS coordinates and radius parameters are used to filter the data source for Macao (latitude = 22.164184, longitude = 113.559866, and radius = 6437 meters). Fig. 2 shows the GPS coordinate center and the control radius of Macao.

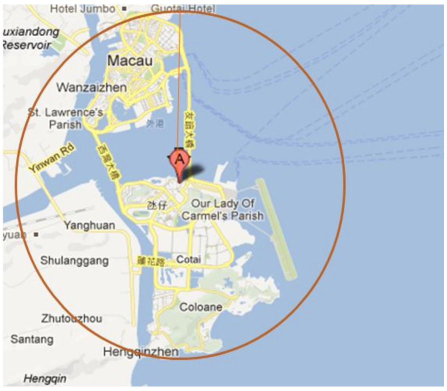


Fig. 2. HTTP Request parameters for Macao.

The returned tweets from the HTTP Response have been stored into a local database for the processing of this work. This work analyzed data from 2013 to 2014. According to the scope of data in the database, most of the Sina Weibo users in Macao are tourists. In 2013, there were 198,506 users who posted tweets in Macao, 9.67% (19,201) were Macao citizens and 90.33% (179,305) were tourists. In 2014, there were 101,049 users who posted tweets in Macao, 12.63% (12,765) were Macao citizens and 87.37% (88, 284) were tourists, as shown in Fig. 3.

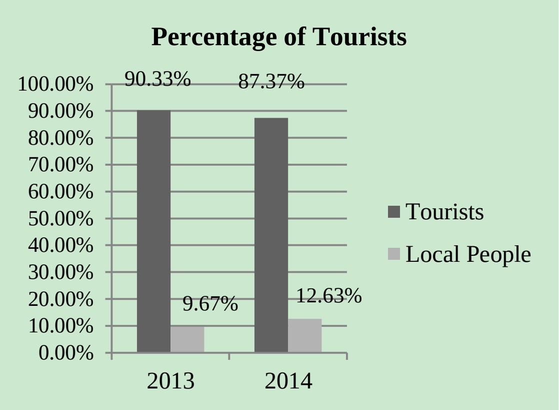


Fig. 3. Percentage of tourists in Macao.

A dataset of 651,490 tweets from 2013 and 331,408 tweets from 2014 is collected from Sina Weibo for this work.

C. Data Preparation and Pre-processing

Since we found that Macao ranked second in China's Feature Tourist City in 2015 [4] [5], the dataset of tweets from Sina Weibo in 2013 and 2014 is analyzed to see how tourist felt about Macao, in terms of the three categories -

Natural Resources, Tourist Goods and History and Culture of the city - grouped together as feature-tourist-city-related posts. The main steps of data pre-processing for our work, as shown in Fig. 4 are as follows:

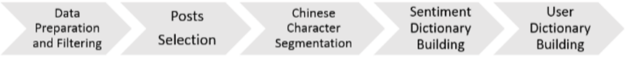

Data Preparation and Filtering
Posts Selection
Chinese Character Segmentation
Sentiment Dictionary Building
User Dictionary Building

Fig. 4. The process of data preparation and pre-processing.

1) Data preparation and filtering

Only tweets related to tourists in the sampled dataset were evaluated and they were treated as bag-of-words since most of the posts were short. Duplicate records were ignored. After filtering, with a reduction of tweets of 56% in 2013 and 70% in 2014, the final dataset of 418,056 tweets from 2013 and 194,880 tweets from 2014 is used to analyze this work. The tweets can be classified as:

- Texts containing positive sentiment, such as happiness or joy.
- Texts containing negative sentiment, such as sadness, anger, or disappointment.
- Texts containing feature-tourist-city-related posts.
- Texts containing feature-tourist-city-related posts with positive sentiment.
- Texts containing feature-tourist-city-related posts with negative sentiment.

2) Posts selection

In this step, we ran classification experiments to validate our dataset, in terms of feature-tourist-city-related posts. Naive Bayes is often used as a baseline in text classification because it is fast and easy to implement [20]. We compared the standard Naive Bayes classifier with the Transformed weight-normalized Complement (TWC) Naive Bayes and found that the standard Na ïve Bayes classifier resulted in slight improvement on all data sets as shown in Table I.

TABLE I: COMPARISON OF TWC AND NA ÏVE BAYES

Data set	Naive Bayes	TWC Naive Bayes
2013 as training set and 2014 as the testing set	87.3%	85.6%
2014 as training set and 2013 as the testing set	85.5%	84.5%

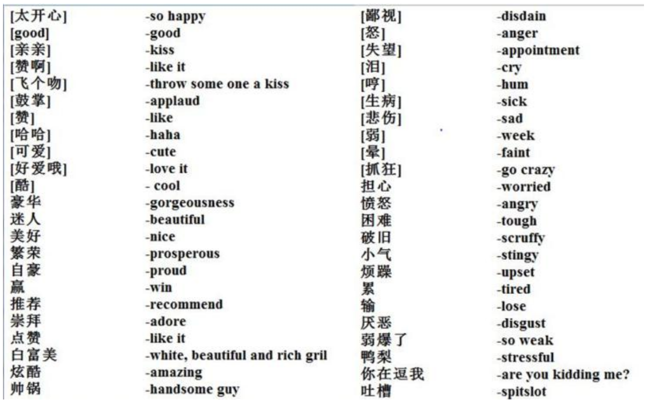
A na ïve Bayes classifier is used in this work in order to classify the Chinese text, this strategy required walking through two successive steps, training and testing. A training document related to topics selected in this work was programmatically generated and manually scanned for accuracy. The application of such a model improved the classification performance, in this case revealed an average performance of 85.5% in 2013 and 87.3% in 2014.

3) Chinese character segmentation/filtering

In this step, a Chinese segmentation program is developed to segment useful words, based on the built-in dictionary in Jieba, using the filtered data from Step (1) because the Chinese characters exported from the tweets of Sina Weibo are sentences and may contain irrelevant characters such as verbs, and other symbols. "Jieba"- (Chinese for "to stutter") is a Chinese text segmentation and according to its author, it is built to be the best Python Chinese word segmentation module [21].

4) Sentiment dictionary building

In this step, the sentiment dictionary in this work is compiled from three different sources, which are Emotional Polarity in Chinese dictionaries from NTUSD [22], sentiment dictionaries from a project called XSimilarity [23], and a hand-annotated dictionary of Sina Weibo icons and network slangs [8]. Examples of positive and negatives words are shown in Fig. 5.



[太开心]	-so happy	[鄙视]	-disdain
[good]	-good	[怒]	-anger
[亲亲]	-kiss	[失望]	-appointment
[赞啊]	-like it	[泪]	-cry
[飞个吻]	-throw some one a kiss	[哼]	-hum
[鼓掌]	-applaud	[生病]	-sick
[赞]	-like	[悲伤]	-sad
[哈哈]	-haha	[弱]	-week
[可爱]	-cute	[晕]	-faint
[好爱哦]	-love it	[抓狂]	-go crazy
[酷]	- cool	担心	-worried
豪华	-gorgeousness	愤怒	-angry
迷人	-beautiful	困难	-tough
美好	-nice	破旧	-scruffy
繁荣	-prosperous	小气	-stingy
自豪	-proud	烦躁	-upset
赢	-win	累	-tired
推荐	-recommend	输	-lose
崇拜	-adore	厌恶	-disgust
点赞	-like it	弱爆了	-so weak
白富美	-white, beautiful and rich gril	鸭梨	-stressful
炫酷	-amazing	你在逗我	-are you kidding me?
帅锅	-handsome guy	吐槽	-spitslot

Fig. 5. Examples of positive and negative words [8].

Emotional Polarity in Chinese dictionary from NTUSD is a very popular sentiment analysis dictionary on the Internet. It classifies words into two different polarities, namely positive and negative. After combining the three sources, our positive sentiment dictionary has 3371 words and our negative sentiment dictionary has 8632 words. The number of positive and negative words in the dataset as compared to our sentiment dictionary are shown in Table II.

TABLE II: NUMBER OF POSITIVE AND NEGATIVE WORDS IN THE DATASET COMPARING TO OUR SENTIMENT DICTIONARY

	Total in sentiment dictionary	Total in dataset			
		2013		2014	
Positive Words	3371	2427	72%	2259	67%
Negative Words	8632	4870	56%	4335	50%

A negation list is constructed based on the word '不' (not) affects the polarity meaning of the phrase, listed in Fig. 6.

其實 呢 份工 我 返到 好 不 開心 (In fact this job I am going to is not too fun)↵

Fig. 6. Sample of word (not) affecting the polarity meaning.

5) User dictionary building

In this step, a user dictionary is constructed manually utilizing the keywords most frequently found in Macao about feature-tourist-city-related areas - Natural Resources, Tourist Goods and History and Culture of the city. These categories are analyzed against our dataset, and our tourists seemed to mention more about tourist goods (see Table III). The observation may be that they are more interested in shopping.

The preparation and pre-processing of data is an important step in the data mining process. The result of segmentation is used for the analysis of the Term Frequency–Inverse Document Frequency while the sentiment dictionary and user dictionary are used to calculate the sentiment analysis in this work.

TABLE III: The Categorization of Feature-Tourist-City

Category	Description [2]	Count	
		2013	2014
1. Natural Resources	They are both a private property and a global common. Destinations that are able to offer travelers access to unique experiences including natural resources have a competitive advantage	1,008	469
		0.2%	0.2%
2. Tourist Goods	Goods that tourists buy will contribute to the Gross Domestic Product (GDP) of that country.	33,750	16,765
		8%	9%
3. History and Culture of the city	Destinations that are able to offer travelers access to unique experience through local culture have a competitive advantage and a basis for generating publicity to attract more awareness, interest, or visitors.	5,681	2,465
		1.4%	1.3%

D. *Data Analysis*

The data analysis is done in two ways: term frequency-inverse document frequency and sentiment analysis.

1) Term frequency–inverse document frequency

We first used the term frequency-inverse document frequency (tf-idf) to initially screen our data. The tf-idf weight is a weight often used in information retrieval and text mining and it is a statistical measure used to evaluate how important a word is to a document in a collection or corpus [24]. It usually assumes the most commonly appearing words in a given document are keywords. Using the data from Step (3) of section III (C) and the built-in tf-idf function in the Jieba segmentation program, we are able to extract the most commonly appearing words as shown in Table IV for 2013 and 2014. The tf-idf function uses the built-in stopwords to process the results.

TABLE IV: Most Commonly Appearing Words Found Using tf-idf Function in Jieba

2013 Chinese	2014 Chinese	English
澳门	澳门	Macao
哈哈	哈哈	Haha
嘻嘻	嘻嘻	Heehee
偷笑	偷笑	giggle
澳門	澳门	Macao
馋嘴	馋嘴	Greedy eater
鼓掌	鼓掌	applaud
花心	花心	Change one’s mind
开心		happy
酒店		hotel
今天	今天	Today
威尼斯人	威尼斯人	Venetian
Macau	Macau	Macao
旅行		travel
	代购	shopping
	拜拜	bye
	分享	share
	面膜	Facial mask

Tourists seem quite happy in Macao, giving the fact that giggle, happy (Haha, Heehee), and applaud are the most commonly appearing words found in our dataset. According to the most commonly appearing words, we can imply to say that tourists travel to Macao to visit Venetian, which correlates with previous finding [8], with Venetian having top tweet quantities in the dataset. They may come to Macao to look for good food or for shopping.

2) Sentiment analysis

Our initial calculation of just counting the frequency of positive words and negative words gave us positive reinforcement in carrying on with the rest of the work. In 2013, we have 283,666 tweets that have positive sentiments and 190,242 tweets that are have negative sentiments. In 2014, we also have 130,193 tweets that have positive sentiments and 87,517 tweets that are have negative sentiments. The frequency table is shown in Table V.

TABLE V: Number of Tweets Showing Positive and Negative Sentiments

	Frequency			
	2013		2014	
Positive words	283,666	68%	130,193	67%
Negative words	190,242	46%	87,517	45%

Our classification method proposed in the previous work [8], is used in this work, with a little modification, which is based on computing the number of the occurrences of positive or negative words in any of the feature-tourist-city-related posts, this will lead to discover the overall satisfaction of the feature-tourist-city topics. We can say that a tweet is a positive, if the total number of the positive words occurrences is more than the total for negative words and vice versa. A tweet is neutral if the score is zero.

Let *w* be each word that matches the sentiment dictionary, *n* (*w*) be the number of negation words that occur before this word. The sentiment score of a tweet is calculated as shown in (1).

$$Score = \sum_{w \in positive}(-1)^{n(w)} * (1) + \sum_{w \in negative}(-1)^{n(w)} * (-1) \qquad (1)$$

Score > 0 → positive sentiment

Score < 0 → negative sentiment

Score = 0 → neutral sentiment

Finally, a program is designed and implemented to capture all possible tweets that contain certain words related to our case study - holding a positive, negative or a neutral opinion or emotion regarding feature-tourist-city-related posts. Initially, the program was seeded with a set of all related keywords, then we did a frequency analysis on the keywords.

IV. Results and Discussion

The results are discussed in this section.

A. Measurement

Overall satisfaction in this work is a measure of how tourists feel about Macao, being one of China's Feature Tourist Cities. In this work, based on the previous work [8], a formula is used to measure the three categories - Natural Resources, Tourist Goods and History and Culture of the city, with three parameters, namely ratio of positive tweets, ratio of negative tweets, and ratio of neutral tweets. The formula is described in (2).

$$Overall\ Satisfaction = \frac{Positive\ ratio + Neutral\ ratio}{Negative\ ratio + Neutral\ ratio} \qquad (2)$$

- If the measure is between 0 and 1, it means Macao brings more negative effects to tourists than positive effects. A smaller value means a worse situation.
- If the measure is approximately equal to 1, it means Macao brings equal amount of positive effects and negative effects to tourists, or Macao does not bring positive effects or negative effects to tourists.
- If the measure is greater than 1, it means Macao brings more positive effects to tourists than negative effects. A larger value means a better situation.

B. Overall Satisfaction of Tourist

This part of the work assesses the overall satisfaction level about the feature-tourist-city-related posts relating to positive and negative sentiment, as shown in Fig. 7. We can see there are definitely more positive sentiments in all categories as well as in both Year 2013 and Year 2014.

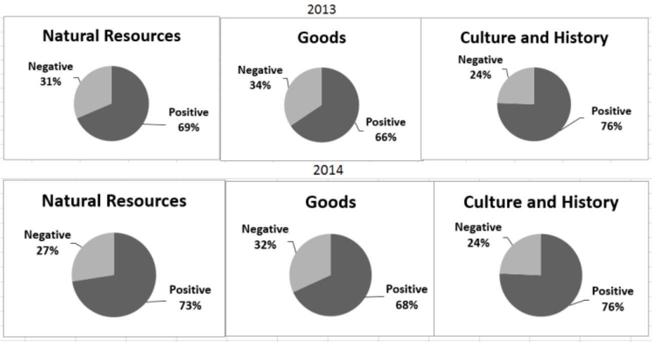


Fig. 7. Positive and Negative sentiments in feature-tourist-city-related posts.

Next the positive, negative, and neutral sentiment for the aggregation of all of the feature-tourist-city-related posts for Year 2013 and Year 2014 is summarized in Table VI.

TABLE VI: POSITIVE, NEGATIVE AND NEUTRAL TWEETS IN 2013 AND 2014

	2013		2014	
Tweets that have scores > 0 (positive)	71,441	17%	35,374	18%
Tweets that have score < 0 (negative)	39,457	9%	17,847	9%
Tweets that are = 0 (neutral)	36,220	9%	16,163	8%

The overall satisfaction of the feature-tourist-city-related posts is calculated from the ratio of positive, negative and neutral tweets of the three categories - Natural Resources, Tourist Goods, Culture and History, using the formula mentioned in (2). It is found that tourists are generally satisfied in Macao in terms of the feature-tourist-city-related posts, with an overall satisfaction of 1.42 in 2013 and an overall satisfaction of 1.52 in 2014. There is an upward trend of 7% in 2014, possibly showing the reason why Macao may be ranked no. 2 in the China's Feature Tourist City Ranking in 2015. A preliminary examination of the text content of the positive and negative tweets shows that the technique used in this work can be used as a sentiment indicator to gauge the tourist satisfaction in Macao.

V. Conclusion

The value of information collected from social networks has been the subject of many studies in different fields. In this paper, we collected social data from Sina Weibo and analyzed our data to see if tourists are satisfied with Macao in terms of three categories - Natural Resources, Tourist Goods and History and Culture of the city. A dataset of 418,056 tweets from 2013 and 194,880 tweets from 2014 is analyzed in this work.

The result indicates that the tourists' overall satisfaction towards Macao is generally positive. In particular, this work comes up with two observations. 1. Tourists come to Macao mainly for shopping, and 2. There is an upward trend of 7% in the overall tourist satisfaction from Year 2013 to Year 2014.

Acknowledgment

This work would not exist without the support and guidance of a group of people. I would like to express my gratitude to Mr. Steven Liu, Mr. Ka Seng Chou, and Mr. Yubin Xiao, for their programming contribution in this work.

References

[1] Internet Live Stats, Internet users. (December 2015). [Online]. Available: http://www.internetlivestats.com/internet-users/

[2] A. Wang. China Internet watch. (December 2015). [Online]. Available: http://www.chinainternetwatch.com/16366/weibo-search-users-insights-2015/

[3] A. Dupeyras and N. MacCallum, "Indicators for Measuring Competitiveness in Tourism: A Guidance Document," *OECD Tourism Papers*, OECD Publishing, 2013.

[4] China's Feature Tourist City. (December 2015). [Online]. Available: http://www.china-citynet.com/yjh/fyphb_show.asp?id=5901

[5] Macao Daily. (December 2015). [Online]. Available: http://www.macaodaily.com/html/2015-12/10/content_1051318.htm

[6] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," *Proceedings of LREC*, 2010.

[7] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in *Proc. the Workshop on Language in Social Media (LSM 2011)*, pp. 30–38, Portland, Oregon, 23 June, 2011.

[8] Y. Luo, P. Lei, and R. Tse, "Evaluating Macao's gaming industry using sentiment analysis on weibo tweets," in *Proc. the 2015 2nd International Conference on Electronic Governance and Open Society: Challenges in Eurasia*, 2015, pp. 139-144.

[9] Y. Li, H. Gao, M. Yang, W. Guan, Ha. Ma, W. Qian, Z. Cao, and X. Yang, "What are Chinese talking about in hot Weibos," *Computing Research Repository* (CoRR) abs/1304.4682, 2013.

[10] H. Yu, G. Sun, and M. Lv, "Users Sleeping Time Analysis based on Micro-blogging Data," in *Proc. 2012 ACM Conference on Ubiquitous Computing*, 2012, pp. 964-968.

[11] J. Chen and J. She, "An analysis of verifications in microblogging social networks - Sina Weibo," presented at the Distributed Computing Systems Workshops (ICDCSW), 32nd International Conference on Computing Systems, 2012.

[12] W. Guan, H. Gao, M. Yang, Y. Li, H. Ma, W. Qian, Z. Cao, and X. Yang "Analyzing user behavior of the micro-blogging website SinaWeibo during hot social events," *Computing Research Repository* (CoRR) abs/1304.3898, 2013.

[13] X. Cui, H. Shi, and X. Yi, "Application of association rule mining theory in Sina Weibo," *Journal of Computer and Communications*, vol. 2, no.1, pp. 19-26, 2014.

[14] Y. Ren, N. Kaji, N. Yoshinaga, and M. Kitsuregawa, "Mining representative posts in Sina Weibo," presented at the 4th Forum on Data Engineering and Information Management, Kobe, Hyogo, March, 2011.

[15] R. Fan, J. Chao, Y. Chen, and K. Xu, "Anger is more Influential than Joy: Sentiment correlation in Weibo," *PLoS ONE*, vol. 9, issue 10, 2014.

[16] F. Yang, X. Yu, Y. Liu, and M. Yang, "Automatic detection of rumor on Sina Weibo," presented at the MDS'12, Beijing, China, August 12, 2012.

[17] K. Wu, S. Yang, and K. Zhu, "False rumors detection on Sina Weibo by propagation structures," presented at the 2015 IEEE 31st International Conference on Data Engineering, Korea, April 13-17, 2015.

[18] T. Tse and E. Zhang, "Analysis of blogs and microblogs: A case study of Chinese bloggers sharing their Hong Kong travel experiences," *Asia Pacific Journal of Tourism Research*, vol. 14, issue 4, pp. 314-329, 2013.

[19] P. Luo and R. Tse, "Research experience of big data analytics: The tools for government - A Case using social network in mining preferences of tourists," in *Proc. the 8th International Conference on Theory and Practice of Electronic Governance*, 2014, pp. 312-315.

[20] J. Rennie, L. Shih, J. Teevan, and D. Karger, "Tackling the poor assumptions of naive Bayes text classifiers," in *Proc. the Twentieth International Conference on Machine Learning*, 2003, pp. 616-623.

[21] Python Package Index – Jieba. (December 2015). [Online]. Available: https://pypi.python.org/pypi/jieba/

[22] Taiwan University. (2013). Emotional Polarity in Chinese dictionaries NTUSD. Datatang. [Online]. Available: http://www.datatang.com/data/44317/

[23] Renmin university of China- iamxiatian/xsimilarity. (December 2015). [Online]. Available: https://github.com/iamxiatian/xsimilarity

[24] Term Frequency–Inverse Document Frequency. (December 20, 2015). [Online]. Available: http://www.tfidf.com/

Rita T. Tse received her bachelor of science degree in math/computer science from UCLA in 1984 and her PhD degree (doctor of philosophy, doing educational technology) from the University of Hull, in England, in 2004.

She is currently the program coordinator of the Computing Program at the Macao Polytechnic Institute. She has been working as an associate professor since 1997 and took up the role of the Program Coordinator since 2004. Her current research interests include Ubiquitous Computing, Urban Sensing, Social Network-based Sensing, Casino Gaming Systems, and Course Design and Development.

Dr. Tse is an associate member of the Institution of Engineering and Technology (IET), and also the Web Chair of Mobicom 2015 (the 21st Annual International Conference on Mobile Computing and Networking). One of her research papers entitled, "Evaluating Macao's Gaming Industry using Sentiment Analysis on Weibo Tweets," was selected as the laureate of the best research papers at the conference "Electronic Governance & Open Society: Challenges in Eurasia". This international conference was organized by the renowned ITMO University in Russia.