Article

A Model to Measure Tourist Preference toward Scenic Spots Based on Social Media Data: A Case of Dapeng in China

Yao Sun [1,2], Hang Ma [1,*] and Edwin H. W. Chan [2]

[1] School of Architecture and Urban Planning, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China; 15903210r@connect.polyu.hk

[2] Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong, China; edwin.chan@polyu.edu.hk

* Correspondence: mahang@hit.edu.cn

Abstract: Research on tourist preference toward different tourism destinations has been a hot topic for decades in the field of tourism development. Tourist preference is mostly measured with small group opinion-based methods through introducing indicator systems in previous studies. In the digital age, e-tourism makes it possible to collect huge volumes of social data produced by tourists from the internet, to establish a new way of measuring tourist preference toward a close group of tourism destinations. This paper introduces a new model using social media data to quantitatively measure the market trend of a group of scenic spots from the angle of tourists' demand, using three attributes: tourist sentiment orientation, present tourist market shares, and potential tourist awareness. Through data mining, cleaning, and analyzing with the framework of Machine Learning, the relative tourist preference toward 34 scenic spots closely located in the Dapeng Peninsula is calculated. The results not only provide a reliable "A-rating" system to gauge the popularity of different scenic spots, but also contribute an innovative measuring model to support scenic spots planning and policy making in the regional context.

Keywords: tourist preference; measuring model; tourism destinations; scenic spots planning; social media data

1. Introduction

As an important industry branch, tourism is a location-based market where different tourism destinations coopetite mutually both in the horizontal and vertical level to occupy the tourist market shares. Because tourism development could provide the most value in stimulating the prosperity of the local economy and improving social welfare, competition rises between different tourism destinations located in the same region, both intentionally and unintentionally, aiming at increasing their popularity among tourists. Hence, it is the ultimate aim of destination competition to gain tourist preference. However, the phenomenon of vicious competition between tourism destinations is very common nowadays in China, which definitely hinders the healthy development of the regional tourism industry. Destinations tend to maximize their own economic benefits, ignoring their actual position in the regional tourist market. Unfortunately, they always imitate and copy another destinations' effective innovations to squeeze the competitors out of the competitive advantage [1]. The phenomenon of cut-throat competition leads to a drastic decline in the number of visitors. For instance, in order to accommodate as many tourists as possible, some scenic spots in the Dapeng peninsula, China, undertake a high-density construction of hotels and restaurants without considering the actual demand of tourists, not only leading to ecological unbalance and resource waste, but also undermining the

overall landscape of tourist attractions. As a result, the current development trend does not comply with the tourist market demand, and many scenic spots have been left bankrupt. Therefore, it is a must to promote tourism development according to tourist preference in the regional context. It is a meaningful topic to measure the actual tourist preference toward different destinations.

E-tourism makes it possible to use a big volume of data on social media (hereafter social media data) to measure the tourist preference toward different destinations. E-tourism offers an effective channel for tourists to arrange travelling plans on the websites, such as selecting tourism destinations, booking hotels and vehicles, and arranging travelling routes [2]. With the advent of e-tourism, the world has experienced the explosion of social media data. Nowadays, social media data not only provides tourists with various travelling information on the demand side, but also inspires tourism destinations to make decisions according to the tourist preferences from the supply side. Using electronic devices, tourists produce a huge volume of data on various social-media websites in China, such as Sina Microblog, Tencent WeChat, and Facebook. Social media data could illustrate the temporal-spatial behavior pattern of tourists, including their geographical mobility, travelling experience, and length of stay in different destinations. Sina Microblog is selected as the data source in this study. Sina Microblog registrants cover nearly 30% of Chinese Internet users, the equivalent of Twitter users in the United States, so the data volume can be guaranteed [3]. Sina Microblog contains a lot of travelling information, because tourists tend to use the website to share their travelling experiences and post comments during their visits. Moreover, Sina Company has opened their data source to the public by developing the Application Programming Interface (API), which guarantees access to the database.

By considering the above reasoning, it reveals that tourism destinations should undertake tourism operations according to tourist preference, which is the compulsory requirement of tourism industry development. Also, social media data is a good choice to measure the tourist preference of destinations. This study tries to measure tourist preference based on social media data. First, we explored the connotation and evaluation system of tourist preference, especially the relationship between tourist preference and destinations' competitive advantages. Second, we elaborated the processing procedure of social media data. Third, we created an original model to measure tourist preference in a quantitative way. Fourth, we applied the above model to the empirical case study of 34 scenic spots in Dapeng Peninsula in order to verify its feasibility. Finally, we recommended some useful tourism planning and management strategies for policy makers based on the measuring results. The novelty of this study lies in the fact that we build an original model to measure tourist preference by applying social media data of e-tourism websites, which lays a solid basis for ameliorating the relevant planning policies for destinations.

2. Literature Review

2.1. The Connotation of Tourist Preference toward Tourism Destinations

Tourist preference conveys a notion of comparison, so it should be discussed on the precondition that the space boundary of the tourism region and the union of whole tourism destinations are well assigned. There are two standards to determine whether different tourism destinations belong to the same tourism region. On one hand, tourism destinations are spatially proximate with high accessibility to each other. On the other hand, similarity of tourism resources exists between different tourism destinations, especially the inherited cultural and natural resources [4].

Tourist preference refers to visitors' perceptions and comments on destinations after an actual visitation, so it can be positive, negative, and neutral. Tourist preference reflects the ability of different destinations to attract tourists and win the tourist market shares by utilizing tourism resources effectively. Besides, tourist preference acts as an important factor of destination competitiveness. According to the Destination Competitiveness Model proposed by Dwyer and Kim in 2003 [5], destination competiveness is influenced by the following four determinants, namely tourism

resources, situational conditions, destination management, and demand conditions. Inherited, created, and supporting tourism resources encompass the various characteristics of a destination that make it attractive to tourists. Situational conditions refer to many types of such factors such as a destination's location, micro and macro environment, and external traffic accessibility. Destination management covers factors that increase the attractiveness of core tourism resources and improve the service quality of the supporting systems [6]. The above three determinants discuss destination competitiveness from the supply side. Meanwhile, demand conditions are mainly determined by tourist preference toward different destinations. Tourist preference can be regarded as the integrated feedbacks of tourism resources, situational conditions, and destination management of a destination. For example, Xu argues that the quality of both core tourism resources and supporting factors together influence tourist preference, because these resources are tourist attractions [7]. As another example, Hearne and Salinas agree that, as well as tourism resources, destination management has an impact on tourist preference [8].

2.2. Evaluating Tourist Preference toward Tourism Destinations

Tourist preference is usually evaluated by the number of visitors, satisfaction degree, and awareness of visitors toward a destination. The reason for measuring tourist preference is that it is very significant for producing visitor satisfaction to match destination tourism products and tourist preference [6]. Nowadays, the majority of current research evaluates tourist preference by adopting small group opinion-based methods such as the analytic hierarchy process (AHP), importance performance analysis (IPA), expert scoring, and questionnaire surveys. The basic data used to seek the scores of every model attribute are usually collected by filed surveys and interviews with partial tourists and relevant stakeholders. For instance, Hearne and Salinas evaluated tourist preference toward two tourist destinations located in Costa Rica based on the data collected by a survey of 171 domestic tourists and 271 foreign tourists [8]. Wang et al. measured tourist preference toward different smart tourist attractions through applying the AHP and IPA approach [9]. Hsu et al. evaluated the tourist preference toward eight destinations in Taiwan via a four-level AHP model depending on the data collected from tourists [10].

To conclude, we cannot deny that the above methods provide a solution to describing tourist preference and achieve meaningful research results. However, these methods are not objective enough to cover all the influential factors, because the scale of the database is limited. With the increasing use of social network platforms on the Internet, it is possible to use a large volume of social media data for the purpose of monitoring the trend of the tourist market. Social media data analysis has greatly improved the accuracy and precision of the evaluation results in the era of e-tourism.

3. Materials and Methods

3.1. Research Area Description

Dapeng Peninsula is located at the southeast of Shenzhen City, Guangdong Province, which is one of the most developed cities in China. It consists of three towns: Kuichong Town in the north, Dapeng Town in the middle, and Nan'ao Town in the south. The peninsula is in fact a district that is surrounded by the gulf from three sides and connected to the mainland in the northwest direction. It borders Huizhou City in the north and Hong Kong Special Administrative Region in the south. Dapeng Peninsula belongs to the subtropical region. The average annual temperature is 22.3 °C, so it is suitable for tourism all the year around. Dapeng Peninsula is an ideal area to develop tourism, whose position makes it a characteristic tourism area based on its unique coastal landscape, long-term history, and good ecological environment [11]. Dapeng Peninsula Official Statistics indicates that the tourism population in 2013 reached three million, and will surpass eight million in 2020 [12]. In the long run, Dapeng Peninsula is planned to be developed into a world-class eco-tourism resort. In Dapeng Peninsula, 34 scenic spots have already become popular tourism destinations (Figure 1).

Figure 1. The geographical location of Dapeng Peninsula in Shenzhen and its administrative divisions.

The space distribution of the 34 tourism destinations is as follows (Figure 2). Nan'ao Town has 17 scenic spots, namely Yangmeikeng Valley, Luzui Villa, Qiniangshan National Geographic Park, Seven-star Yacht Club, Dapeng Yacht Club, Langqi Yacht Club, Qiqi Family Paradise, Riding Club, Judiaosha Sandbeach, Dongshan Pier, Dongchong Sandbeach, Xichong Sandbeach, Shuitousha Sandbeach, Lover Island, Moon Sandbeach, Bantianyun Village, and Nan'ao Pier. Ten scenic spots are located in Dapeng Town, namely Jinshawan Sandbeach, Xiantouling Prehistoric Habitation, Jiaochangwei Sandbeach, Pengcheng Village, Dongshan Temple, Paiya Mountain, Dayawan Nuclear Plant, Wangmuwei Village, Guanyin Mountain, and Longyan Temple. The scenic spots in Kuichong Town are seated in the western seaside, referring to Guanhu Sandbeach, Shayuchong Sandbeach, Old Headquarters of the Dongjiang Army, Tuyang Sandbeach, Xiyong Sandbeach, Rose Sandbeach, and Baguang Scenic Spot. The core tourism resources of Dapeng Peninsula are located in the 34 scenic spots.

Tourism resources in the 34 scenic spots are different. According to Dwyer and Kim [5], tourism resources can be classified into three types, namely inherited cultural and natural resources, created resources, and supporting resources. Inherited cultural and natural resources refer to cultural heritages and natural landscape with tourist attractions. Created resources are man-made entertainment projects or activities. Supporting factors and resources refer to systems such as the accommodation system, telecommunication system, financial institutions, and currency exchange facilities. As for Dapeng Peninsula, the most attractive resources are the inherited coastal sandbeach and mountain landscape with high forest coverage. Moreover, the local inherited cultural resources and man-made created resources also play an important role in regional tourism development. The classification of tourism resources in Dapeng Peninsula is listed in Table 1. Among them, 18 scenic spots attract tourists based on inherited coastal natural resources, such as sandbeach and coastal valley. Four scenic spots center on inherited mountain natural resources. Six scenic spots are famous for inherited historical and cultural resources. The other six scenic spots rely on created resources to develop tourism, such as yacht clubs and family paradise.

Figure 2. The spatial location of tourism destinations classified by core tourism resources.

Table 1. The classification of tourism resources of scenic spots in Dapeng Peninsula.

Core Tourism Resources	Scenic Spots
Inherited coastal natural resource	Yangmeikeng Valley, Luzui Villa, Judiaosha Sandbeach, Dongshan Pier, Dongchong Sandbeach, Xichong Sandbeach, Shuitousha Sandbeach, Lover Island, Moon Sandbeach, Nan'ao Pier, Jinshawan Sandbeach, Jiaochangwei Sandbeach, Guanhu Sandbeach, Shayuchong Sandbeach, Tuyang Sandbeach, Xiyong Sandbeach, Rose Sandbeach and Baguang Scenic Spot
Inherited mountain natural resource	Qiniangshan National Geographic Park, Paiya Mountain, Guanyin Mountain, Old Headquarters of the Dongjiang Army
Inherited historical and cultural resource	Xiantouling Prehistoric Habitation, Pengcheng Village, Wangmuwei Village, Dongshan Temple, Longyan Temple, Bantianyun Village
Created resources	Seven-star Yacht Club, Dapeng Yacht Club, Langqi Yacht Club, Qiqi Family Paradise, Riding Club, Dayawan Nuclear Plant

3.2. Social Media Data

3.2.1. Data Mining

In order to play an important role in guiding the decision-making process, Sina Company has opened its microblog data source to the public by developing the Internet link port, API. By connecting the API link at the website of "http://open.weibo.com/", more than twenty categories of data sources can be accessed [13]. Through the API, Sina Company opens a certain volume of real-time microblogs posted by the registrants all around the world. They can be downloaded in any Internet development environment with user authorization. After obtaining the user authorization, python programming language is applied to create the data mining tool, known as the web crawler [14]. Therefore, the data mining is a process of interdisciplinary cooperation between computer science and tourism planning. In this study, the selected data source contains the information of users' identity, content of microblogs, release time and location, and the number being reposted, liked, and commented on. For the purpose of improving the accuracy of tourist preference of scenic spots, the volume of data should be sufficient. In this study, it takes 89 days to conduct data mining through the API, resulting in a total of 2000 million pieces of microblogs.

3.2.2. Data Cleaning

As for the mined 2000 million pieces of microblogs, they should be filtered and cleaned to delete the interference data which is not relevant to this study. Data cleaning criteria mainly contain two aspects, including recognizing the targeted tourism destinations and identifying tourist behaviors. On one hand, the microblogs linked to targeted scenic spots are filtered by semantic recognition [15].

The purpose of semantic recognition is selecting out the microblogs that contain the words and expressions listed in the text corpus. Taking Dapeng Peninsula as an example, the microblogs containing the name of any one of the 34 scenic spots should be selected out. In addition, the microblogs released in the spatial boundary of the 34 scenic spots should also be chosen. In order to capture the targeted microblogs, the text corpus was built first. In the text corpus, the formal and informal names of the 34 scenic spots were fully listed. Especially, scenic spots whose names are usually written by mistake need particular attention. For instance, the Chinese name of Xichong Sandbeach, "西涌沙滩", is usually written as "西冲沙滩". Qiniang Mountain National Geological Park is also texted in the microblogs as Qiniang Mountain Park or Nan'ao Geological Park. Therefore, the wrong writing forms and informal writing forms of scenic spots should be listed in the text corpus aiming at collecting valid microblogs that are as complete as possible. After building the complete text corpus, the initially mined 2000 million microblogs were computed one by one through text comparison with the text corpus. In this way, 10,132 pieces of microblogs were preliminarily selected. Then, measures were taken to avoid 1218 overlapping names of places which are not located in Dapeng Peninsula. At this stage, we achieved 8914 microblogs.

On the other hand, the 8914 microblogs filtered by semantic recognition were further cleaned to identify tourist behaviors piece by piece. Finally, there were 6819 microblogs filtered, which were edited by Chinese characters and qualified for the following data analysis. The collected 6819 microblogs were released from January 2010 to December 2016, so the time range of data is seven years. On this basis, 6819 cleaned microblogs were classified and linked to every scenic spot that they described. Every microblog contains the information of users' identity, content of microblogs, release time and location, and the number being reposted, liked, and commented on.

3.3. Model to Measure Tourist Preference of Scenic Spots

3.3.1. Tourist Sentiment Orientation

After classifying microblogs according to the name of scenic spots annually, the number of microblogs linked to every destination and the times of being reposted, liked, and commented are

calculated. In this study, tourist sentiment orientation toward different destinations is analyzed by applying Chinese Sentiment Analysis to the text content of microblogs [16]. Chinese Sentiment Analysis is based on the framework of Machine Learning in computer science. Machine Learning aims to make computers gain artificial intelligence that endows them the capabilities to make decisions [17]. This research applies the framework of Machine Learning to empower computers with the artificial intelligence to judge the sentiment orientation of microblog content linked to every one of the specified destinations [18].

Firstly, the text content of microblogs is coded in the form of word vectors that can be recognized by a computer with the means of Word2Vec. Then, textual content is translated into vector-address code. Secondly, training corpus data is input in the Long Short-Term Memory Model for the purpose of making the computer memorize the sentiment orientation of different structures of word vectors by changing the parameters of the Long Short-Term Memory Model [19]. The training corpus data refers to the vector-address code, which is originally written in Chinese characters and verifies their sentiment orientation [20]. Thirdly, the translated vector-address code of microblog content is input in the Long Short-Term Memory Model. A computer uses artificial intelligence to determine the sentiment orientation of microblogs by comparing the vector structure with the training corpus data.

Taking Dapeng Peninsula as an example, the training corpus data base consisted of 500 million vector-address codes that had verified the sentiment orientation in advance. It took 24 days to input the training corpus data in the Long Short-Term Memory Model to ensure that the computer had the intelligence to memorize the sentiment orientation of different structures of word vectors by modifying the parameters. Then, 100 million training corpus data was again input in the modified model to check the accuracy of this model. The accuracy degree was checked as 95.54%. On this basis, the 6819 prepared microblogs linked to different destinations were input in the model to test the tourist sentiment orientation of each destination.

The sentiment orientation was valued from 0 to 1. When the value equals 0.5, the tourist sentiment orientation of the corresponding tourism destination is neutral. When the value is below 0.5, the tourist sentiment orientation tends to be negative. When the value is above 0.5, the tourist sentiment orientation tends to be positive.

3.3.2. Present Tourist Market Shares

Based on the number of microblogs linked to every one of the scenic spots, the present market shares of every destination can be calculated. In a narrow sense, the tourist market shares are equal to the competitiveness [21]. Although the number of microblogs are not equal to the absolute value of actual tourist flows, it can infer the comparative value of different destinations because the tourist sentiment orientation is valued from 0 to 1. In order to standardize all the three dimensions, the present market shares should also be normalized by a value from 0 to 1. The present tourist market shares are defined as follows:

$$A_i = B_i / B_{max} \tag{1}$$

In this formula, A_i means the present market shares of certain scenic spots, B_i represents the actual number of microblogs linked to a specific scenic spot, and B_{max} means the maximal number of microblogs.

3.3.3. Potential Tourist Awareness

The number of microblogs being reposted, liked, and commented on linked to different destinations can reflect the potential tourist awareness. Because tourist awareness revealed by reposting, liking, and commenting on microblogs is different, the value weight of the three aspects should be distinguished. In this study, expert scoring is selected to determine the attribute weight. In total, 20 experts were invited to give their answers on the expert questionnaire. In the questionnaire, experts were invited to give scores on the weight of reposts, likes, and comments. In total, 16 effective

questionnaires were acquired, and the effective return ratio was 80%. Among these experts, five experts were from the field of tourism management, five experts studied social media data technology, and six experts explored the strategies of scenic spot planning. According to the statistics, the weights of reposts, comments, and likes are set as 1/2, 1/3, and 1/6. The potential tourist awareness is defined as follows:

$$C_i = (D_i/2 + E_i/3 + F_i/6)/C_{max} \qquad (2)$$

In the formula, C_i means the potential tourist awareness; D_i, E_i, and F_i represent the actual number of reposts, comments, and likes, respectively; and C_{max} means the maximal potential tourist awareness.

3.3.4. Integrated Model

The tourist preference of scenic spots can be measured by the above three dimensions, namely tourist sentiment orientation, present tourist market shares, and potential tourist awareness. In order to endow the above three dimensions with reasonable weights, the method of expert scoring is adopted. After calculating the average of scores marked by 16 experts in the expert questionnaire, present tourist market shares, tourist sentiment orientation, and potential tourist awareness are endowed the weight value of 0.75, 0.15, and 0.1, respectively. Present tourist market shares reflect the attraction of every tourism destination to tourists, so as a decisive factor of tourist preference, the tourist sentiment orientation represents the tourist comments on the current service quality of different scenic spots. Potential tourist awareness predicts the underlying attraction of scenic spots. The tourism destinations' competitiveness is measured as follows:

$$G_i = 100 \times (0.75 \times A_i + 0.15 \times Q_i + 0.1 \times C_i) \qquad (3)$$

In the formula, G_i means the tourist preference of scenic spots, and A_i, Q_i, and C_i represent the present market shares, tourist sentiment orientation, and potential tourist awareness of certain scenic spots, respectively.

To conclude, the technical route to measure the tourist preference of scenic spots based on social media data is as follows (Figure 3). This technical route reveals the procedure of building the original model and can be adopted as a useful tool to measure the tourist preference of any close groups of tourism destinations.
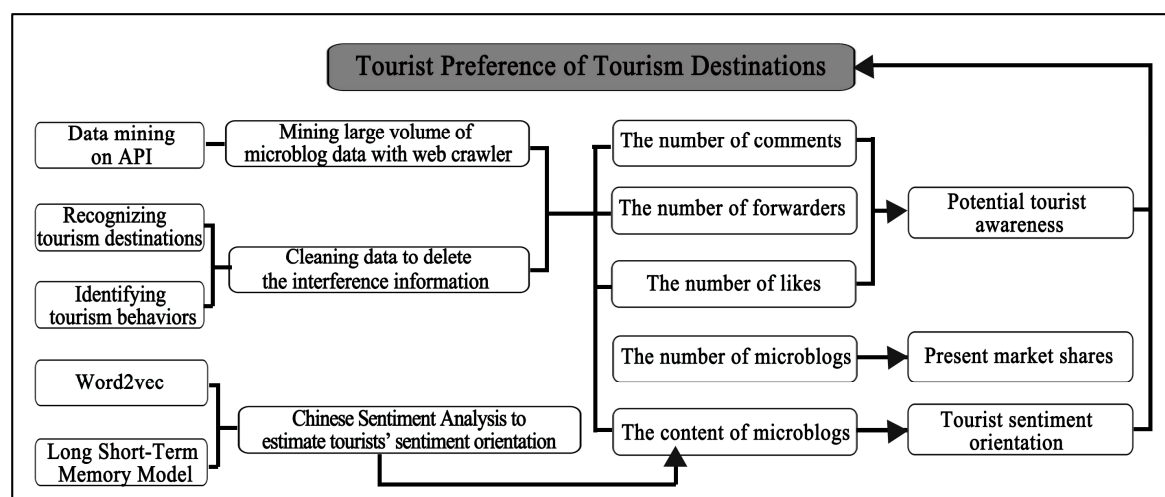

Figure 3. Technical route of creating a model to measure the tourist preference of scenic spots.

4. Result and Validity Analysis

4.1. Result

Taking Bantianyun Village as an example, the tourist preference of this village was measured with the above methods. Bantianyun Village is one of the 34 scenic spots in Dapeng Peninsula, known as the most beautiful traditional village in Guangdong Province (Figure 4). It is located in the west of Nan'so Town, surrounded by mountains. In the village, Hakka folk buildings are well conserved and the original rural landscape is attractive. Within the 6819 pieces of microblogs, there were four microblogs linked to Bantianyun Village, which were all posted on Sina Microblog in 2016. Based on the statistics, the number of reposts, comments, and likes are 0, 5, and 11, respectively. The original microblogs content edited in Chinese related to Bantianyun Village is translated in English (Table 2). In Dapeng Peninsula, the maximal number of microblogs is 414, linked to the Yangmeikeng Valley. According to the Formula (1), the present tourist market shares of Bantianyun Village are equal to 0.0097. Depending on Formula (2), the maximal potential tourist awareness is calculated as 1773. Hence, the potential tourist awareness of Bantianyun Village is equal to 0.001973684. With the method of Chinese Sentiment Analysis based on Machine Learning Technology, the tourist sentiment orientation of the microblog text content is computed as 0.67, reflecting that the Bantianyun Village provides tourists with a positive travelling experience. Based on Formula (3), the tourist preference of Bantianyun Village is 10.79.

Figure 4. The landscape of Bantianyun Village.

Table 2. Social Media Data of Sina Microblogs linked to Bantianyun Village.

User ID	English Translation of Weibo Posts	Posting Time	Repost	Comment	Like
Vegetable 66_13351	I published the blog named 25 October 2015 Bantianyun Village in Shenzhen, what a quiet rural village.	16 June 2016—15:30	0	0	0
Quietly brilliant 91785	Raining scene in Bantianyun Village.	3 June 2016—20:12	0	0	0
Yirui_gogo	#Shenzhen Bantianyun# I am watching the seafront landscape while hiking. There are a few houses. The folk houses are constructed using black bricks. Roads are paved with slabstones. What a paradise.	15 May 2016—23:09	0	0	2
SailorHZ	I Encountered Bantianyun Village while cycling. The village is located in Nan'ao Town, and it is one of the most beautiful villages with good ecological environment. Bantianyun Village is situated in half-way up the hill (the altitude is 428 m). It is said Bantianyun Village is the highest ancient village in Shenzhen. I must climb to the gate of Damaotian Reservoir, and there is a perverted slope. #Shenzhen Bantianyun#	9 April 2016—19:49	0	5	9

Similarly, the tourist preference of all the 34 scenic spots in Dapeng Peninsula can be generated (Table 3). According to the tourist preference in total, 34 scenic spots are ranked in descending order. Compared with the other 33 tourism destinations, the tourist preference of Bantianyun Village is not evident, and the present market shares are relatively low. It is less than two years since Bantianyun Village undertook tourism operation, so the majority of tourists are not aware of the destination. Besides, the inadequate tourist service facilities and disorganized management are other important reasons for the tourism depression in Bantianyun Village.

Table 3. Result of annual tourist preference of the 34 scenic spots in Dapeng Peninsula.

Rank	Tourism Destinations	2016	2015	2014	2013	2012	2011	2010	Total
1	Yangmeikeng Valley	87	33	31	47	70	0	0	268
2	Xichong Sandbeach	45	34	29	61	61	0	0	230
3	Jiaochangwei Sandbeach	54	96	34	20	12	0	0	216
4	Pengcheng Village	76	22	15	21	21	13	0	168
5	Qiniangshan National Geographic Park	33	22	36	27	19	9	8	154
6	Jinshawan Sandbeach	11	11	10	27	43	44	0	146
7	Dongchong Sandbeach	11	13	12	23	30	24	14	127
8	Guanhu Sandbeach	18	25	20	21	20	15	0	119
9	Rose Sandbeach	30	24	22	39	0	0	0	115
10	Dongshan Temple	16	21	19	22	21	15	0	114
11	Xiyong Sandbeach	15	12	9	16	16	21	12	101
12	Moon Sandbeach	11	13	11	19	23	20	0	97
13	Seven-star Yacht Club	17	15	18	21	23	0	0	94
14	Old Headquarters of the Dongjiang Army	14	13	14	19	15	19	0	94
15	Dapeng Yacht Club	18	14	16	15	16	12	0	91
16	Baguang Scenic Spot	16	22	19	33	0	0	0	90
17	Shayuchong Sandbeach	11	11	10	3	16	14	8	73
18	Guanyin Mountain	13	15	17	22	0	0	0	67
19	Tuyang Sandbeach	11	14	11	15	12	0	0	63
20	Xiantouling Prehistoric Habitation	11	9	10	7	13	9	0	59
21	Paiya Mountain	10	17	14	15	0	0	0	56
22	LuzuiVilla	16	12	10	14	0	0	0	52
23	Qiqi Family Paradise	14	11	10	0	0	0	0	35
24	Riding Club	0	11	0	12	11	0	0	34
25	Longyan Temper	0	0	0	11	9	11	0	31
26	Lover Island	10	9	11	0	0	0	0	30
27	Shuitousha Sandbeach	7	10	9	0	0	0	0	26
28	Dongshan Pier	18	0	0	0	0	0	0	18
29	Judiaosha Sandbeach	17	0	0	0	0	0	0	17
30	Wangmeiwei Village	0	0	8	0	0	9	0	17
31	Dayawan Nuclear Plant	14	0	0	0	0	0	0	14
32	Langqi Yacht Club	13	0	0	0	0	0	0	13
33	Nan'so Pier	11	0	0	0	0	0	0	11
34	Bantianyun Village	11	0	0	0	0	0	0	11

4.2. Validity Analysis

According to the result measured by the model, 34 scenic spots in Dapeng Peninsula are classified into five ratings based on their tourist preference in total (Table 4). Among them, Yangmeikeng Valley, Xichong Sandbeach, and Jiaochangwei Sandbeach are set as the 5A destinations with the greatest popularity among tourists, acting as the core tourism attractions in the region.

Table 4. A-ratings of the 34 scenic spots in Dapeng Peninsula.

A-Ratings	Scenic Spots	Tourist Preference	Grading
A	Qiqi Family Paradise, Riding Club, Longyan Temple, Lover Island, Shuitousha Sandbeach, Judiaosha Sandbeach, Wangmuwei Village, Bantianyun Village, Dayangwan Nuclear Plant, Langqi Yacht club, Nan'ao Pier	(0, 50)	Weak
AA	Moon Sandbeach, Seven-star Yacht Club, Old Headquarters of the Dongjiang Army, Dapeng Yacht Club, Baguang Scenic Spot, Shayuchong Sandbeach, Guanyin Mountain, Longyan Temple, Tuyang Sandbeach, Xiantouling Prehistoric Habitation, Paiya Mountain, LuzuiVilla	(50, 100)	
AAA	Jinshawan Sandbeach, Dongchong Sandbeach, Guanhu Sandbeach, Rose Sandbeach, Dongshan Temple, Xiyong Sandbeach	(100, 150)	
AAAA	Pengcheng Village, Qiniangshan National Geographic Park	(150, 200)	
AAAAA	Yangmeikeng Valley, Xichong Sandbeach, Jiaochangwei Sandbeach	(200, +∞)	Strong

In order to verify the reliability of the result measured by the model, the tourist preference of the 34 scenic spots in Dapeng Peninsula was double-checked by 30 employees from the government of Dapeng District. The selected respondents must satisfy the following two conditions to ensure that they have a comprehensive understanding of the tourism development of Dapeng Peninsula. First, they are all direct and indirect decision makers of local tourism development. Second, they have kept themselves on the job in tourism management for more than three years.

In the survey, the result of A-ratings based on tourist preference was presented to the 30 respondents. They were asked to judge if the five grades of the 34 scenic spots conform to reality according to their experience. Their decisions were made based on the actual number of tourist receipts, tourist complaints, and brand name awareness linked to different scenic spots. The above three standards correspond to the three attributes in the model, namely present tourist market shares, tourist sentiment orientation, and potential tourist awareness. After a one-hour discussion on the A-ratings result, 27 respondents agreed that the A-ratings are reasonable and can reflect the actual degree of popularity of scenic spots. The level of matching in comparison to the result of our measuring model is 90%, and thus our model is verified as valid.

5. Conclusions and Suggestions

5.1. Conclusions

Under the background of e-tourism popularity, this study further taps into the value of social media data in tourism development. We designed an original technical route to illustrate the procedure of data mining, data cleaning, and data analyzing. First, data mining was conducted on the API of Sina Microblog with a coded web crawler. Second, we proposed the methods of data cleaning by recognizing the targeted tourism destinations and identifying tourist behaviors. Third, we introduced

Word2Vec and the Long Short-Term Memory Model to conduct data analysis. We believe this study creates a manipulable tool to connect social media data with the smart tourism development of scenic spots.

Following the technical route, we create an original model to measure the tourist preference of scenic spots located in the same region by introducing three attributes, namely present market shares, tourist sentiment orientation, and potential tourist awareness. Tourist sentiment orientation is measured by applying Chinese Sentiment Analysis based on the text content of microblogs. Present market shares are calculated based on the number of microblogs linked to every scenic spot. Potential tourist awareness is reflected by measuring the number of microblogs being forwarded, liked, and commented on. This model was then applied to the empirical case study of 34 scenic spots in Dapeng Peninsula. The results indicated the relative tourist preference of every scenic spots.

To conclude, this study focuses on applying social media data to quantitatively measure the tourist preference of scenic spots located in the same region. The result is of great value for policy makers associated with tourism planning and management. This model could be used as a valuable guideline to guide future scenic spot planning in the regional context.

5.2. Suggestions

The model is an effective tool to help monitor the operation situation of every scenic spot located in the same region. From the demand side, we can know about the feedback of tourists. By conducting Chinese Sentiment Analysis on the text content of microblogs, tourists' travelling experience toward destinations can be assessed with a high accuracy. In this sense, the preferences of tourists are well captured. On this basis, policies should be made to follow the travelling demand of tourists by highlighting the spots with positive feedback and improving the ones with negative feedback. From the supply side, it is necessary for managers to stay informed of the capacity of scenic spots. For example, present tourist market shares can infer the spatial distribution of visitors and the number of tourist receptions. If the tourist number exceeds the ecological capacity, measures should be taken to evacuate passenger flows to destinations with a sufficient reception ability.

Moreover, the model provides an effective solution to grade the A-ratings of different scenic spots. For smart regional tourism development, hierarchical planning strategies of scenic spots should be made according to their tourist preference. Destinations with a strong tourist preference should sustain their development policies by sticking to successful experiences, and promote the tourism prosperity of other destinations by undertaking model-driven development mode. Destinations with weak tourist preference should explore the potential of tourism resources, and take advantage of the external economy effect of those first-rated destinations by sharing tourist market and supplementing tourist services.

Author Contributions: Yao Sun conducted data analysis, designed the model, and drafted the main text of this paper. Hang Ma and Edwin H. W. Chan conceived the study, supervised the process of manuscript development, and supported the research funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tolstad, H.K. Development of rural-tourism experiences through networking: An example from Gudbrandsdalen, Norway. *Nor. J. Geogr.* **2014**, *2*, 111–120. [CrossRef]
2. Miller, R.L.; Frank, B.C. *The Legal and E-Commerce Environment Today*, 7th ed.; South-Western College/West: Mason, OH, USA, 2012; pp. 7–21.

3. Sina Microblog Data Center. Annual Development Report on Sina Microblog Registrants 2016. Available online: http://data.weibo.com/report/reportdetail?id=346 (accessed on 4 September 2017).

4. Wu, J. Defining tourism cluster and identifying the boundary of regional tourism. In Proceedings of the Academic Proceedings of Chinese Geography Society, Hengyang, China, 29 April 2009.

5. Dwyer, L.; Kim, C. Destination competitiveness: Determinants and Indicators. *Curr. Issues Tour.* **2003**, *5*, 369–413. [CrossRef]

6. Gomezelj, D.O.; Mihalič, T. Destination competitiveness—Applying different models, the case of slovenia. *Tour. Manag.* **2008**, *2*, 294–307. [CrossRef]

7. Xu, S. *The Basic Theory on Evaluating Regional Tourism Competitiveness*; Northeast Normal University: Changchun, China, 2006.

8. Hearne, R.R.; Salinas, Z.M. The use of choice experiments in the analysis of tourist preferences for ecotourism development in Costa Rica. *J. Environ. Manag.* **2002**, *65*, 153–163. [CrossRef]

9. Xia, W.; Xiang, L.; Feng, Z.; Zhang, J.H. How smart is your tourist attraction?: Measuring tourist preferences of smart tourism attractions via a FCEM-AHP and IPA approach. *Tour. Manag.* **2016**, *54*, 309–320.

10. Tzukuang, H.; Yifan, T.; Wu, H.H. The preference analysis for tourist choice of destination: A case study of Taiwan. *Tour. Manag.* **2009**, *30*, 288–297.

11. Shenzhen Urban Planning and Land Resource Research Center. Shenzhen Industrial Space Layout Planning (2011–2020). Available online: http://ghzx2014.tianv.net/product-fruit-i_8304.htm (accessed on 4 September 2017).

12. Urban Planning, Land and Resource Commission of Shenzhen Municipality. *The Conversation and Development Planning for Dapeng New District (2013–2020)*; Shenzhen Municipal Government: Shenzhen, China, 2013.

13. Lian, J.; Zhou, X.; Cao, W. Data mining solutions on Sina Microblog. *J. Tsinghua Univ.* **2011**, *10*, 1300–1305.

14. Mirzal, A. Design and implementation of a simple web search engine. *Comput. Sci.* **2011**, *5*, 53–60.

15. Liu, J.; Zhou, Y.; Gao, F. Identifying sensitive subjects on the Internet based on semantic recognition. *J. Shenzhen Inf. Technol. Coll.* **2011**, *3*, 33–37.

16. Peng, H.; Cambria, E.; Hussain, A. A review of sentiment analysis research in Chinese language. *Cogn. Comput.* **2017**, *4*, 1–13. [CrossRef]

17. Mitchell, M.T. *Machine Learning*, 1st ed.; MIT Press: Boston, MA, USA, 2003.

18. Michalski, S.R.; Carbonell, G.J.; Mitchell, M.T. *Machine Learning: An Artificial Intelligence Approach*, 1st ed.; Tioga Publishing Co.: Palo Alto, CA, USA, 1984; Volume II.

19. Burgess, N.; Hitch, G.J. A revised model of short-term memory and long-term learning of verbal sequences. *J. Mem. Lang.* **2006**, *4*, 627–652. [CrossRef]

20. Zhang, Z.; Weninger, F.; Wöllmer, M.; Schuller, B. Unsupervised learning in cross-corpus acoustic emotion recognition. In Proceedings of the 2011 IEEE Workshop on Automatic Speech Recognition & Understanding, Waikoloa, HI, USA, 11–15 December 2011; pp. 523–528.

21. Durbarry, R.; Sinclair, M.T. Market shares analysis: The case of French tourism demand. *Ann. Tour. Res.* **2003**, *4*, 927–941. [CrossRef]