Journal of Computational Science & Engineering 30 (2017) 853-858

Available at http://www.asocse.org ISSN 1710-4068

Application of Data Mining Technology in Tourism Image Creation

Zhenlei Yu

Information Ministry of Library, Qilu University of Technology Jinan 250100, China

Abstract: **With advances in computer technology, adopting data mining to analyze and process the assessment information of tourists in travel destinations has become an important way of doing research on tourism image. In this paper, taking Mount Tai scenic spot as an example, with the support of ROST Content Mining software, the author employs the method of content analysis to search for the tourists' online assessment information, and probes into the tourists' perception of the tourism image of Mount Tai, in the hope of providing a basis for the enhancement of tourism image.**

Keywords: Data mining technology; Content analysis method; Tourism image

Received 3 July 2017, Revised 23 July 2017, Accepted 25 July 2017

*Corresponding Author: Zhenlei Yu

1. Introduction

In the "Internet + " background, the way of getting access to and sharing information has changed greatly, and the change is particularly reflected in tourists. Based on their own personal experiences in travel destinations, tourists can post their general and abstract emotional evaluation on twitter, wechat, and various tourism websites. The evaluation information will have an influence on the future tourist consumption in scenic spots. Positive evaluation can enhance the image of scenic spots, attracting more tourists, and negative evaluation will have a negative impact on the tourism image of scenic spots. Therefore, the study of tourists' evaluation of the scenic spots has become one of the hot research topics in the field of tourism.

With the rapid development of computer technology such as big data and cloud computing, to analyze and review the tourist' online assessment information has become an significant approach to the study on tourism image of travel destinations. The research on tourism image abroad started in 1970s and has achieved a lot of valuable research results. For example, Joseph and Chenting explained the behavioral foundation of the perception of tourism image, and established a comprehensive model integrating destination image, self-congruity, and travel behavior[1]. Domestic research in this area started late. Huang Zhenfang and Li Xiang (2002) believed that the study on measurement of travel destination image should be based on Structured Approach and Unstructured Approach[2]. In this paper, based on the previous study, the author takes Mount Tai as the research object, and adopts content analysis method to dig into the tourists' online assessment information with the help of ROST Content Mining software. By doing this, the author intends to explore the tourists' perception of the tourism image of Mount Tai, and provide a foundation for the enhancement of tourism image.

2. Research Method

2.1. Research object

Mount Tai is the world's cultural and natural heritage, attracting a lot of visitors. Because of its great popularity, it receives comprehensive assessment information from the tourists on the Internet. The assessment information is representative and typical. Therefore, the author of this paper takes Mount Tai as the object of study.

2.2. Research approach

Taking the tourist's online reviews as the sample, the author first uses the software of ROST Content Mining[3] to sort out the words and select the high-frequency words, having a preliminary understanding of the general characteristics of Mount Tai tourism image. Then the method of content analysis[4] is adopted to build the category entries of analysis, to encode the sample and verify the reliability of the coding results, so as to obtain the frequency distribution of the primary and secondary category entries of tourism image. In this way, the author deepens the understanding of the tourism image of Mount Tai. At last, the author evaluates the tourism image of Mount Tai through comparing the evaluation frequency and the average score of the primary and secondary category entries of tourism image attributes.

2.3. Research Method and Procedure

2.3.1. Sample Selection and Processing

The data come from the tourists' online reviews provided by China's largest online travel service provider, Ctrip. The statistical data came out on March 10th 2016. From October 2014 to February 2016, there are altogether 380 reviews from the tourists, and except for the 21 useless reviews including the duplicate reviews, reviews have nothing to do with the topic, and inquiries about travel information, 359 reviews are left to be valid.

2.3.2. Analysis of the High-Frequency Characteristic Words

2017 © The American Computational Science Society. All rights reserved.

In this paper, the software of ROST Content Mining is used for word frequency analysis. First, the words related to Mount Tai are included in the custom dictionary, and they are sorted out so as to get the word frequency statistics. Then after screening, the meaningful characteristic words with high frequency are selected for the analysis of the theme of Mount Tai tourism image. In this way, the author has a general knowledge of the tourists' perception of Mount Tai tourism image.

2.3.3. Content Analysis Method

Categories of analysis are built to define the identification rules and coding requirements. Based on the analysis results of high-frequency words obtained by ROST, the tourists' assessment information, and the actual situation of Mount Tai, the category entries of tourism image attributes are established, consisting of six primary category entries and eighteen secondary category entries. They are presented in Table 1 as follows.

Table 1 Category Entries of Tourism Image Attributes

Primary category entries	Secondary category entries	Primary category entries	Secondary category entries
Landscape	1.Overall landscape 2.Architectural landscape 3.Mountain landscape 4.Plant landscape 5.Overall atmosphere	Accommodation	14.Accommodation facilities
Environment	6.Local prices 7.Local residents 8.Infrastructure 9.Climatic conditions 10.Air quality 11.Environmental sanitation	Transportation	15.Transportation for entering the scenic spot 16.Roads inside the scenic spot
Catering	12.Overall catering 13.Special snacks	Touring	17.Entire scenic area 18.Each attraction

Content analysis unit is the actual computation object during the process of encoding. In this paper, the review data of every tourist is encoded as a unit of analysis, and four coders are arranged to code the 359 analysis unit. For the secondary category entries, if involved in the analysis unit, they are coded as "1", otherwise "0". With regard to the primary category entries, if one or more secondary category entries belonging to the primary category entries are touched upon in the analysis unit, they are coded as "1", otherwise "0".

After the coding is completed, the statistical results are obtained to test the reliability[5]. The formula of the reliability discriminated interactively among the four coders is as follows[6]:

$$R = \frac{n \times K}{1 + (n-1) \times K} \quad (1)$$

In the formula, R represents the coefficient of the interactive discrimination, n is the number of coders, and K is the average degree of mutual agreement between two of the coders.The average degree of mutual agreement between the two coders is calculated as:

$$K_{AB} = \frac{2M_{AB}}{N_A + N_B} \quad (2)$$

In the formula, M_{AB} stands for the number of the analysis units that get the identical results from the two coders, N_A is the number of the analysis units that coded by the coder A, and N_B is the number of the analysis units that coded by the coder B. The average degree of mutual agreement between every two of the four coders is K_{AB}, K_{AC}, K_{AD}, K_{BC}, K_{BD}, and K_{CD}. The average degree of mutual agreement among the four coders is calculated as:

$$K_{ABCD} = \frac{K_{AB} + K_{AC} + K_{AD} + K_{BC} + K_{BD} + K_{CD}}{4} \quad (3)$$

The interactive discrimination reliability can be obtained based on the above three formulas (see table 2). If the value of reliability is higher than 0.80, it means the results are acceptable; if the value is higher than 0.90, the results more credible. In this paper, the reliability value of each of the primary and secondary category entry is greater than 0.90, at a high level, signifying the coding results are valid.

2.4. Evaluation of Tourism Image

Tourism image evaluation include the tourists' evaluation of the overall tourism image of Mount Tai and the evaluation of secondary category entries of tourism image attributes. The evaluation value of the overall tourism image is directly obtained based on the tourists' online assessment information; the scores of

2017 © The American Computational Science Society. All rights reserved.

the secondary category entries of tourism image attributes are determined by the tourists' evaluation. The results of both the two parts are represented by the five-point Richter scale. The numbers 1 to 5 respectively represent "very bad, bad, fair, good, and very good"[7]. As for the reliability of the secondary category entries of tourism image attributes, the test method is the same as that of content analysis. Upon testing, the interactive discrimination reliability value of each secondary category entry is above 0.90, proving the evaluation results are credible.

Table 2 The interactive discrimination reliability between coders

Tourism image attributes		Interactive discrimination reliability		Tourism image attributes		Interactive discrimination reliability	
Primary category entries	Secondary category entries	Primary category entries	Secondary category entries	Primary category entries	Secondary category entries	Primary category entries	Secondary category entries
Landscape	1.Overall landscape	0.937	0.908	Catering	12.Overall catering	0.980	0.995
	2.Architectural landscape		0.980		13.Special snacks		0.994
	3.Mountain landscape		0.975	Accommodation	14.Accommodation facilities	0.970	0.997
	4.Plant landscape		0.933	Transportation	15.Transportation for entering the scenic spot	0.952	0.947
Environment	5.Overall atmosphere	0.925	0.935				0.980
	6.Local prices		0.964				
	7.Local residents		0.989				
	8.Infrastructure		0.950		16.Roads inside the scenic spot		
	9.Climatic conditions		0.917				
	10.Air quality		0.943				
	11.Environmental sanitation		0.970	Touring	17.Entire scenic area	0.921	0.940
					18.Each attraction		0.985

3. Results and Analysis

3.1. High-frequency characteristic words

Table 3 High-frequency characteristic words about Mount Tai tourism image

Characteristic words (frequency)	Part of speech	Characteristic words (frequency)	Part of speech	Characteristic words (frequency)	Part of speech	Characteristic words (frequency)	Part of speech
Mount Tai(221)	Noun	The Eighteen Bends (Shibapan)(36)	Noun	Experience(17)	Verb	Sea of cloud(10)	Noun
Sunrise(126)	Noun	Ropeway(31)	Noun	Weather(16)	Noun	Tai'an City(10)	Noun
Come down the Mountain(78)	Verb	The Supreme of Five Sacred Mountain(29)	Noun	Overcoat(15)	Noun	Regretful(9)	Adjective
Climb the Mountain(77)	Verb	Ticket(29)	Noun	Exercise(15)	Verb	Hike(9)	Verb
Peak of the Mountain(76)	Noun	Suggest(29)	Verb	Morning(14)	Noun	Happy(9)	Adjective

2017 © The American Computational Science Society. All rights reserved.

Hour(68)	Noun	Queue up(27)	Verb	Footstep(14)	Noun	sacrifices to heaven(9)	Noun
Landscape (58)	Noun	Attraction(26)	Noun	Physical strength(14)	Noun	Grand ceremony(9)	Noun
Convenience(54)	Noun	Rain(24)	Verb	thin pancake (14)	Noun	Magnificent(8)	Adjecti ve
The Southern Heavenly Gate (Nantianmen)(50)	Noun	A panoramic view across mountains(23)	Adjecti ve	Arrive(12)	Verb	Extremely ling(8)	Noun
Night(45)	Noun	Peak of the Heavenly Emperor (Yuhuangding)(21)	Noun	Route(12)	Noun	Sense of achievement(8)	Noun
Splendid(43)	Verb	Be worth watching(19)	Adjecti ve	Reach the peak(12)	Verb	Hotel(8)	Noun
Mountaineer(40)	Verb	Take a bus(19)	Verb	Accommodation(1 2)	Noun	Culture(8)	Noun
Scenery(40)	Noun	Take a rest(18)	Verb	Huge crowds of people(12)	Adjecti ve	Wonderland (8)	Noun
Cable car(37)	Noun	Spectacular(17)	Adjecti ve	Delicious(11)	Adjecti ve	Shocking(7)	Adjecti ve
Feel(36)	Verb	Be a feast for the eyes(17)	Adjecti ve	Beautiful (10)	Adjecti ve	Stone inscription(7)	Noun

This paper finally obtains 60 high-frequency words related to the tourism image of Mount Tai (see table 3). It can be seen from the table that the high-frequency words are mainly nouns, verbs, and adjectives. Among them nouns refer to the names of the travel destinations and buildings, a total of 34, the largest number, accounting for 56.67%; verbs refer to the words describing the tourists' mountain climbing experiences, a total of 15, accounting for 25%; adjectives refer to the words describing the environment, landscape features, and snacks of Mount Tai scenic spot, a total of 11 with the least number, accounting for 18.33%.

First, it can be seen from the 10 most frequently used words that "Mount Tai", as the name of a travel destination, is mentioned the most by tourists, and "sunrise" stands for the important landscape of Mount Tai. As a matter of fact, "watching sunrise at the peak of Mount Tai" is one of the landscapes in which many tourists are interested. Second, based on the high-frequency characteristic words, the themes of Mount Tai tourism image can be grouped into the unique geographical conditions, unique cultural environment and atmosphere, beautiful architectural landscape and tourist attractions, and delicious snacks.

3.2. Frequency Statistics of Tourists Image Attributes

In general, the tourism image attribute mentioned a lot by tourists refers to the one that makes a deep impression on the tourists, reflecting the prominent part of tourist's perception of destination image[8]. Therefore, the frequency statistics of the primary and secondary category entries of tourism image attributes is of great importance for exploring the structure and content of Mount Tai tourism image. Table 4 shows that the primary category entries of tourism image attributes are mentioned 752 times by tourists among which landscape 243 times, environment 232 times, catering 45 times, accommodation 55 times, transportation 90 times, and touring 87 times, accounting for 67.9% , 64.6%, 12.5%, 15.3%, 25.1% and 24.1% of the total respectively. The results indicate that the landscape, environment, transportation, and touring make the deepest impression on tourists, and they are the most prominent elements of Mount Tai tourism image. The frequency statistics of the secondary category entries of tourism image attributes reveal that the eighteen secondary category entries are mentioned 1853 times. Ranking the top five are overall atmosphere with 349 times, each attraction 343 times, architectural landscape 201 times, roads inside the scenic spot 167 times, and mountain landscape156 times, accounting for 97.2%, 95.5%, 60.0%, 46.5%, and 43.4% respectively.

Table 4 Frequency statistics of the primary and secondary category entries of tourists image attributes

Tourism image attributes	The number/percentage of comments	Tourism image attributes	The number/percentage of comments

2017 © The American Computational Science Society. All rights reserved.

Primary category entries	Secondary category entries	Primary category entries	Secondary category entries	Primary category entries	Secondary category entries	Primary category entries	Secondary category entries
Landscape	1.Overall landscape	243/67.9%	90/25.1%	Catering	12.Overall catering	45/12.5%	53/14.8%
	2.Architectural landscape		201/60.0%		13.Special snacks		62/17.3%
	3.Mountain landscape		156/43.4%	Accommodation	14.Accommodation facilities	55/15.3%	80/22.3%
	4.Plant landscape		92/25.6%	Transportation	15.Transportation for entering the scenic spot	90/25.1%	53/14.8%
Environment	5.Overall atmosphere	232/64.6%	349/97.2%				167/46.5%
	6.Local prices		13/3.6%		16.Roads inside the scenic spot		
	7.Local esidents		17/4.7%				
	8.Infrastructure		30/8.4%				
	9.Climatic conditions		35/9.7%	Touring	17.Entire scenic area	142/39.6%	52/14.5%
	10.Air quality		15/4.2%		18.Each attraction		343/95.5%
	11.Environmental sanitation		45/12.5%				

3.3. Evaluation of Tourism Image

SPSS19.0 software is used for the statistical analysis of the evaluation results. Firstly, 359 tourists make an evaluation of Mount Tai tourism image, and among them, 248 tourists think it is “very good”, 84 tourist “good”, 22 tourists “fair”, and the number of tourists think it is “bad” and “very bad” is 3 and 2 respectively. The average value of the evaluation is 4.596, indicating that the tourists’ evaluation of Mount Tai tourism image is high with most of them giving a score of “4” and above.

Secondly, the evaluation results of the eighteen secondary category entries of tourism image attributes are analyzed, and the results are ranked from high to low, as shown in Exhibit 1 and Exhibit 2. Generally speaking, the more the secondary category entry is evaluated, the more attention it draws from the tourists; the higher the score is, the more satisfactory the tourists are about the category entry[9]. As a consequence, the evaluation results of the secondary category entries are analyzed based on the correspondence between the evaluation times and evaluation scores.

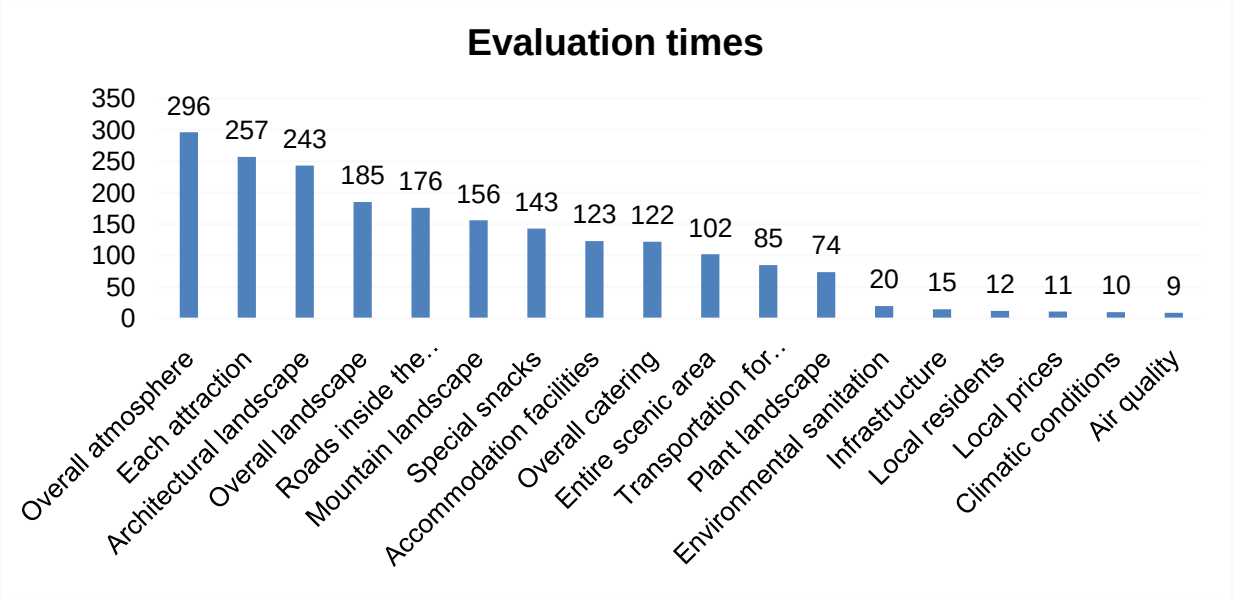

Exhibit 1. The bar chart of the evaluation times of secondary category entries.

2017 © The American Computational Science Society. All rights reserved.

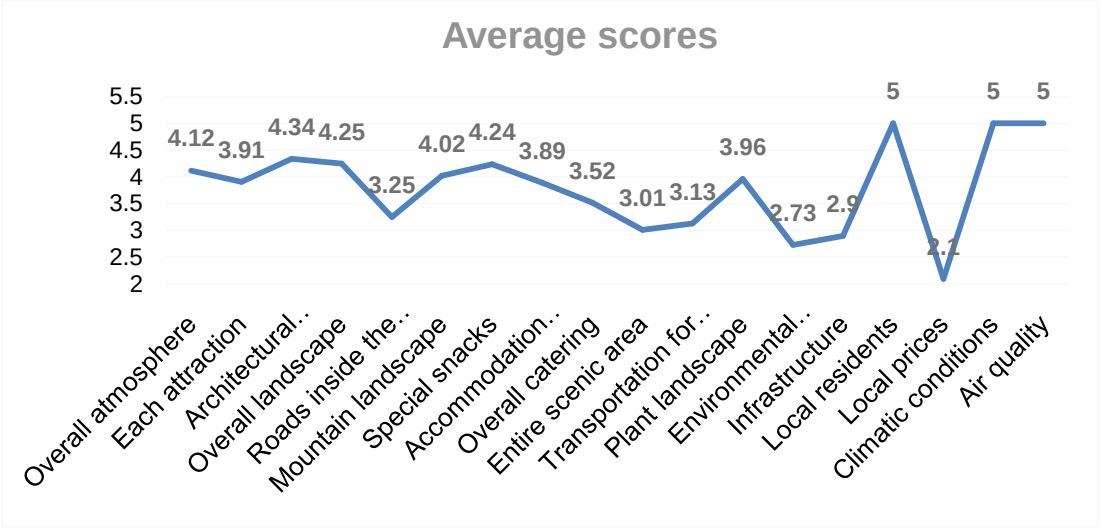

Exhibit 2. The line chart of the average scores of secondary category entries.

It can be seen from the above exhibits that tourists pay more attention to the tourism image attributes of Mount Tai such as the overall atmosphere, each attraction, architectural landscape, overall landscape, roads inside the scenic spot, mountain landscape, special snacks, accommodation facilities, overall catering, entire scenic area, transportation for entering the scenic spot, and plant landscape. All of the secondary category entries are evaluated for more than 50 times, and the average scores are all above "fair", especially for "overall atmosphere", "architectural landscape", "mountain landscape", and "special snacks", the evaluation result is between "good" and "very good". In the future, therefore, these secondary category entries should be maintained so as to further enhance the satisfaction of tourists.

Besides, there are four secondary category entries obtaining an evaluation of "fair". The tourists are dissatisfied with the roads inside Mount Tai scenic spot. The main reason is that the mountain road is featured by plenty of narrow footsteps, making it difficult for the tourists to climb up, and the toughest part is the Eighteen Bends (Shibapan) which fatigues the tourists and makes their enthusiastic decline. As for "overall catering", the reason for the poor evaluation may be that the food is overpriced and does not taste good. The reason for the poor evaluation of "entire scenic area" probably is that most of the historic attractions are seriously damaged and the bad scenic environment caused by an excess of tourists. With regard to "transportation for entering the scenic spot", it obtains a poor evaluation mostly because there are too many tourists and the transportation costs are high.

4. Conclusion

Taking Mount Tai as the example, with the support of ROST Content Mining software, the present paper analyzes the tourists' online assessment data by adopting the method of content analysis, and explores the attributes as well as features of Mount Tai tourism image. Based on the category entries obtained through content analysis, the author evaluates the tourism image of Mount Tai, providing a foundation for improving Mount Tai tourism image. However, the data dug up in this paper is not enough for the research on tourism image. In future study, more attention should be paid to the travel guides, travel notes, blogs, and reviews on tourism websites. As far as data analysis is concerned, more data mining methods should be adopted to improve the reliability of data mining and theoretical research. In terms of data statistics, the methods of correlation analysis and cluster analysis should be employed to make the data analysis more scientific.

References

[1] Z.F. Huang, X. Li, Y.X. Gao. Nankai Business Review, 3 (2002): 69-73.
[2] S. Joseph, S. Chenting. Journal of Travel Research, 5 (2000): 340-352.
[3] L. Xiao, L.M. Zhao. Tourism Tribune, 24 (2009):75-81.
[4] R.Y. Sun. Modern Information, 1(2005):2-6.
[5] W. Wang. Chaidamu Kaifa Yanjiu, 6 (2006):47-49.
[6] O.R. Holsti. Content Analysis for The Social Sciences and Humanities[M].Don Mills: Addison-Wesley Publishing Company, 1969.
[7] S.K. Ma, W. Wang. Science &Technical in formation of Gansu, 1 (2007):68-151.
[8] J. Gao, Y.G. Zhang, D.Q. Zhuang. Consumption economy, 25 (2009): 62-65.
[9] W. Xiong, J.H. Xu. Journal Beijing International Studies University, 11 (2010):57-67.

2017 © The American Computational Science Society. All rights reserved.