Gender difference on destination image and travel options: An exploratory text-mining study

Rui Wang
School of Economics and Management
Beihang University
Beijing, China
wrhfut@126.com

Jin-Xing Hao
School of Economics and Management
Beihang University
Beijing, China
Hao@buaa.edu.cn

Abstract*—Destination image has gained increasingly attention of tourism researchers. However, little attention has been devoted to study the difference in perception of destination image among different demographic groups using the big data analytics in general, the text mining approach in particular. Therefore, in this paper, we develop a method based on Latent Dirichlet Allocation (LDA) to extract cognitive destination image from travel blogs. Furthermore, we explore the differences in cognitive image and travel options among different gender groups. Through experiments, we extract 26 topics by LDA and find that the female has significantly preferences to natural landscape and rural scenery than the male. Meanwhile, the male prefers historical sites than the female. In terms of the travel options, we find the female usually travel with travel agency, while the male travel by self-drive. Our study contributes to the tourism planning, tourism positioning and marketing for different target groups.

Keywords*—Tourism destination image; Gender groups; LDA;

I. INTRODUCTION

Research on destination image has always been a hot topic in the field of service management. With people's living standards increase, more and more people choose to enjoy life through tourism. At the same time, with the progress of society, the demand of everyone is personalized. Different people prefer different tourism destinations and different travel options. However, as far as extant research is concerned, most of scholars have focused on study destination image using classical survey approaches. Seldom research has devoted to examine differences in perception of destination image among different gender groups using the big data analytics in general, the text mining approach in particular. Thus, in this research we focus on the impact of gender on the tourist cognitive image of tourism destinations and the relations between gender and travel options using an exploratory text-mining method. The findings are helpful to formulate personalized marketing strategies for tourism destination and market segmentation.

II. LITERATURE REVIEW

Research on destination image began in the 1970s. The tourism destination image is generally considered to be the overall understanding and evaluation of the tourist destination by tourists and potential tourists. It is a synthesis of the destination society, politics, economy, life, culture and tourism development. Baloglu et al. [1] has developed a formation model of tourism destination image. They believe that destination image is affected by personal factors and stimulus factors. Gallarza and Saura [2] present a conceptual model of tourism destination image, which is consists of four feature, which are complex, multiple, relativistic and dynamic nature. Beerli and Martin [3] elaborate that tourism destination image include cognitive image, affective image and conative image. Meanwhile, their views are widely accepted by most scholars. Kelly and Mackay [4] propose that factors influence the image of tourism destination include income, familiarity, gender, etc., and they also incorporate gender differences in their model. Based on this research, several researchers have studied the relations between gender groups and tourism. For example, Frew and Shaw [5] find that the male are more likely to prefer some sporting games than the female. Xie and Bao [6] indicate that the female usually buy souvenirs after traveling an attraction, while the male do not. Kim et al. [7] illustrate that there are differences between the male and the female in searching online information. They draw to the conclusions that the male in general do not use external information, while the female will look for implicate information in website with positive attitudes. Abubakar et al.[8] explore the difference of revisit destination between the male and the female. They find the impact of electronic word-of-mouth on the trust of destination is about 1.2 times higher for the male, while the female are 2.3 times higher. Elias et al. [9] observe that the male spend more time on travel and activities outside than the female, and the reason underlying this phenomenon may be attributed to more males hold a driving license than females. The study of López-Mosquera [10] indicates that the female usually consider others suggestions more than the male when buy and pay for something. Figueroa-Domecq[11] proposed that male and female have different demands and expectations during travel. Wang et al. [12] confirm that the effects of travel motivation on cognitive image, advertising on cognitive image and tourist expectation were significantly stronger for the male than the female. In summary, prior research supports that the male and the female differ in many aspects. Tourists of different genders have different concerns in the process of travel. The female think that travel is a process of relaxation, an important way enjoying life and broaden their horizons. The male is more concerned with attractions and increasing friendships. The reason why this phenomenon is that there are differences between male and female in tourism motives, tourism behaviors, tourism activities and tourism experiences [13].

Previous studies have provided us with some supports, but there are still exist several limitations. First, although the existing research has done a lot of studied between gender and tourism, seldom research focuses on the cognitive image of different gender groups. Second, for the measurements of destination image measurement, prior research relies on traditional survey method instead of recent big data analytics, which is increasingly important in the digital economy. Therefore, this research attempts to explore the relations among cognitive image, travel options and gender groups

978-1-5386-5178-0/18/$31.00 ©2018 IEEE

using a text mining method. As a result, our research targets to solve follow questions:

1. How to extract cognitive image of a destination expressed in online travel blogs?

2. How different gender perceive cognitive image of a destination?

3. How different gender prefer different travel options?

Thus, the contribution of this research is that we develop a method based on Latent Dirichlet Allocation (LDA) to extract cognitive destination image from travel blogs, and explore the relations among cognitive image, travel options and gender groups. The result will be helpful to the formulation of marketing strategies for tourism destination according to the different target groups.

III. METHOD

In this study, we design a three-step process to extract cognitive image of a destination expressed in online travel blogs. The first step is the data collection. It includes the content of travel blogs, the gender of author and the travel options of author. The second step is cognitive image of tourism destination extraction from travel blogs using a Latent Dirichlet Allocation (LDA model) based text mining method. The third step is data analysis. We analyze differences of gender among the cognitive image of the tourism destination and the travel options between different genders.

A. Data Source

Ctrip travel is one of the largest platform to share tourism information. Tourists can share their travel experience by write travel blogs. Furthermore, travel blogs can be collected with a lower cost. Thus, we developed a web crawler to obtain travel blogs. For statistical convenience, we selected 10 cities of China, and a total of 10,000 travel blogs for experiments after remove unsuitable data. For each travel blog, we obtained that tourist or writer's gender and travel options. When tourist select travel options, each tourist could select more than one choice. The travel options contain semi-guided tour, independent travel, with travel agency, food, humanities, train, petty bourgeoisie, photography, save money, weekend tour, hiking, self-drive, riding, luxury, poor travel, shopping and cruise ship. Independent travel refers to all plans in the trip are made by self even meet trouble things. Semi-guided tour refers to a trip with travel agency. The cost includes only transportation and accommodation expense; tourists should make decision about the rest expenses. Petty bourgeoisie refers to tourists is emphasis lifestyle and spiritual enjoyment during the journey.

B. Data pre-processing

Before the topic extraction, we preprocessed the data. It contained removing invalid characters, word segmentation, eliminating stop words and word tokenization. The file of stop words could be updated according to each result of experiment. We implemented data pre-processing in the python programming environment.

C. Latent Dirichlet Allocation

The LDA model is a document topic generation model [14]. It gives the topic of each document in the datasets with a distribution of probabilities. In LDA model, each document can be represented a probability distribution of topics, and each topic can be represented a probability distribution of many words. Based on this principle, we can identify the most prominent topics in each travel blog. Furthermore, we can name each cognitive topic according to the word distribution in each topic. LDA is an unsupervised model that can handle high-dimensional data. In particular, it is suitable for processing a large number of text datasets.

IV. EXPERIMENT

In this section, we reported the general process of our experiment. After data collection and pre-processing, we used the LDA model to extract topic. Finally, we analyzed the relations among topic of cognitive image, travel options and gender groups. All experiment steps were conducted in the Windows environment.

A. Sampling

TABLE I. CHARACTERISTICS OF TRAVEL BLOGS IN SAMPLE (N=10,000)

		Freq	**%**
City	Beijing	1,200	12%
	Dalian	1,000	10%
	Guilin	1,000	10%
	Jinzhong	1,000	10%
	Kunming	1,000	10%
	Luoyang	1,000	10%
	Nanjing	1,000	10%
	Shenzhen	1,000	10%
	Shijiazhuang	800	8%
	Weihai	1,000	10%
Gender	Male	3,998	39.98 %
	Female	6,002	60.02 %
About travel options	Male	559	5.59 %
	Female	600	6.00 %
	Undisclosed	8,841	88.41 %

B. Topic extraction

In this research, we used the LDA package of python to run LDA model. In the program, we inputted the text data which has been preprocessed. Then, we set the parameters of LDA model, including iteration, alpha, eta and topics. Usually, beta=0.05 and alpha=50/topics. According to the perplexity index[15], we selected 26 topics as optimal number of topic.

V. RESULTS

In this section, we report results of our experiment. Firstly, we list results of topics of destination image extraction by LDA model in Table II and Fig. 1. Secondly, we report the relations between gender and topics of cognitive image in Fig. 2. Thirdly, we illustrate the connection between gender and each travel choice in Fig. 3.

A. Topics of tourism destination image

Table II lists the 26 most important cognitive image topic of tourism destination from 10,000 travel blogs by LDA. The corresponding frequency for each topic is given in Fig. 1.

In terms of the frequency, we can observe that seaside, traditional building, religious culture, natural landscape, historical sites, tourism service, memorial building, park and rural scenery are more mentioned by tourists than other topics in travel blogs.

TABLE II. A LIST OF EXTRACTED TOPICS

Topic name	Topic name	Topic name
Plateau landscape	Plant	Tourism service
Photography	Religious culture	Accommodation
Seaside	Resorts	Ancient building
Entertainment	Ethnic customs	Memorial building
Revolutionary sites	Natural landscape	Art
Transportation	Alley life	Park
Traditional building	Snow landscape	Diet cooking
Modern building	Leisure	Rural scenery
Catering services	Historical sites	

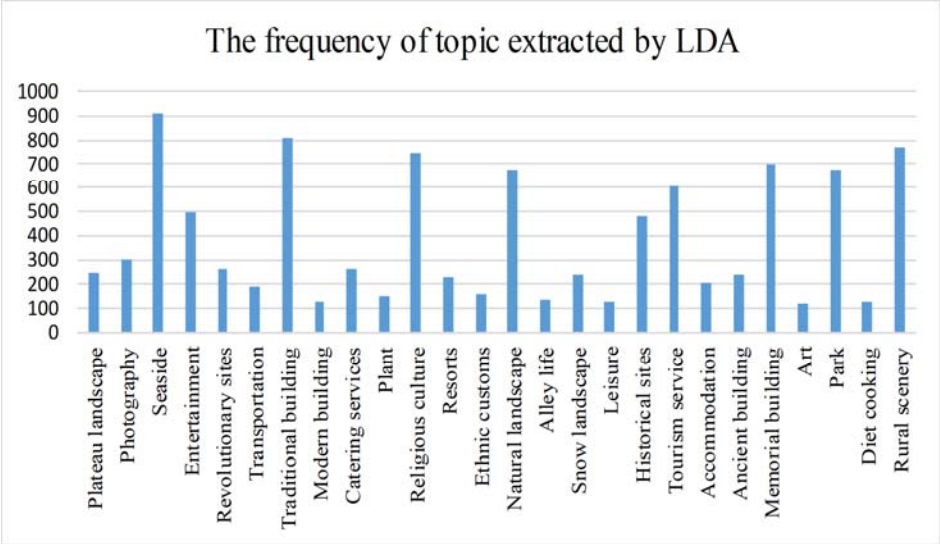

Fig. 1 The frequency of topic extracted by LDA

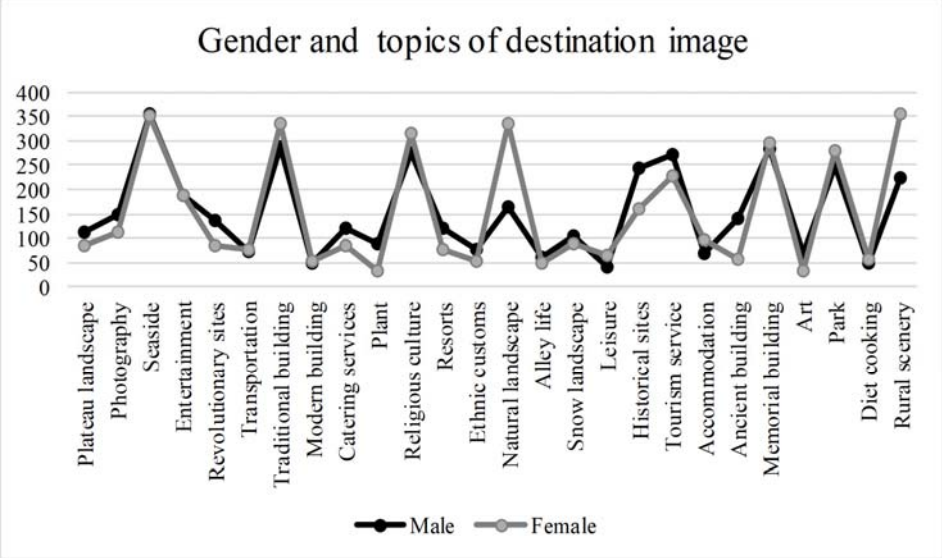

Fig. 2 The relations between gender and topics of cognitive image

B. The relations between gender and topics of cognitive image

Considering that the proportion of the male is 39.98% and the female is 60.02%, which may influence the result, we selected 65.6% samples (3,938 samples) from the female tourists by random sampling. Fig. 2 combines the results of both gender and topics of cognitive image into a single graph. The female has significantly mentioned of natural landscape (t=-8.041, p=0.000) and rural scenery (t=-5.720,

p=0.000) than the male. Meanwhile, the male may prefer revolutionary sites (t=3.357, p=0.001), historical sites (t=4.165, p=0.000) and plateau land-scape (t=2.132, p=0.033) than the female.

C. *The relations between gender and travel options*

Fig.3 presents the proportion of the male and the female in each travel choice. We can find that the female usually travels with travel agency (t=-2.704, p=0.007), while the male is travel by self-drive (t=4.502, p=0.000).

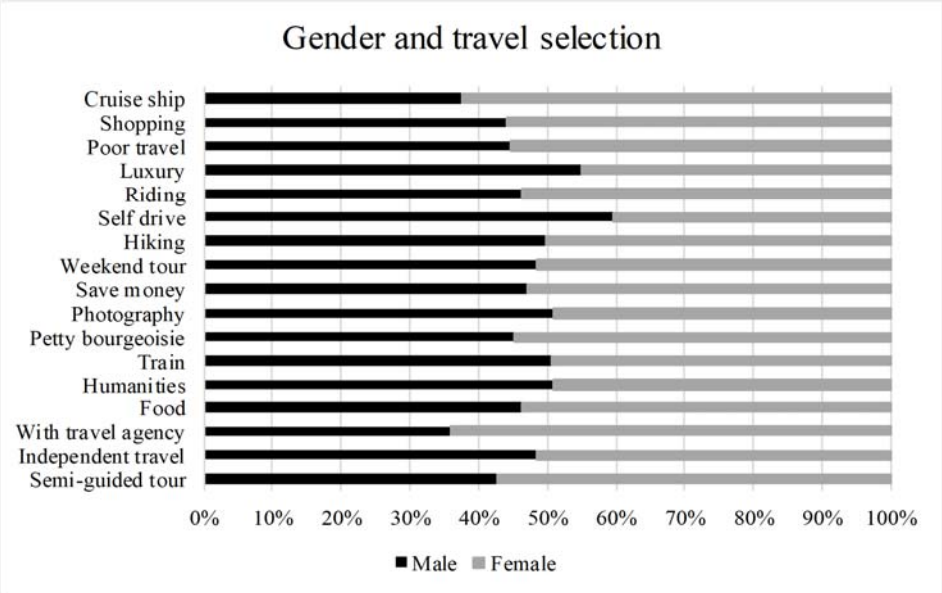

Fig. 3 The relations between gender and travel options

VI. DISCUSSION

In this research, topics of tourism cognitive image destination are extracted from 10,000 travel blogs using LDA model, we explored the relations among cognitive image, travel options and gender groups. According to our dataset, the proportion of the female is more than the male. Maybe the female prefer to write travel blogs than the male, or the female prefer to travel more than the male. It is contrast to prior research[16]. The reason for this phenomena may be that with the development the pace of life, the male has more pressure of working or living than the female, and the female with more free time to traveling by convenient transportation. Some interesting findings can be observed from Fig. 2 and Fig. 3. The female prefers natural landscape than the male, and they also prefer to enjoying nature and relaxed life. However, the male prefer history more than the female. In particular, the male prefer to plateau landscape more than the female, and it may be that the male are stronger than the female, and they do not worry about high altitude reaction and other adverse reactions. About travel options, the female is always travel with travel agency, while the male prefer to drive by themselves. This may be due to the security problems of the female traveling alone. For the male, they do not generally consider safe issues, and overall more males hold a driving license than females. This support the findings of prior studies[9]. This study also has important practical implications. In natural landscape and historical sites, destination management organizations (DMOs) can develop destination marketing strategies based on the characteristics of the female and the male. For example, in cities with great natural landscapes, relevant management person can provide advertising and promotion for the female target groups. Furthermore, in cities with a long historical culture and sites, the relevant tourism departments can hold some interesting activities to attract the male target groups. Travel agency can offer some discounts for the female tourists, and make travel routes that meet the demand of the female in slack seasons. For some self-driving tourists, DMOs can use bundling sale strategy for tickets and parking fees. Meanwhile, they can offer some tips of driving tour for them. Through some marketing strategies, the topics of destination cognitive image plays an important role in the target market.

VII. LIMITATION AND FUTURE WORK

This study has a few limitations. First, we don't analyze the changes of the tourism destination cognitive image in different gender with time. But, in reality, people's preferences are changing with the development of society and time. Second, Ctrip.com does not publish the information of tourists' age. Thus, the relations between different age groups and destination cognitive image has not been studied. Third, in order to convenience of statistics, we just use 10,000 travel blogs in China, some latent cognitive has not been found based on limited data. In the future work, we will collect more travel blogs in depth (e.g., age and profession) and width (e.g., other country and other language), and get more interesting findings.

ACKNOWLEDGMENT

Jin-Xing Hao is the corresponding author of this paper. We would like to acknowledge the supports from projects funded by National Science Foundation of China (Nos. 71471011, 71531001), and Fundamental Research Funds for the Central Universities of BeiHang University.

REFERENCES

[1] S. Baloglu and K. W. McCleary, "A model of destination image formation," Annals of Tourism Research, vol. 26, pp. 868-897, 1999.

[2] M. G. Gallarza and I. G. Saura, "Destination image towards a conceptual framework," Annals of Tourism Research, vol. 29, pp. 56-78, 2002.

[3] A. Beerli and J. D. Martín, "Factors influencing destination image," Annals of Tourism Research, vol. 31, pp. 657-681, 2004.

[4] D. R. F. Kelly J.Mackay "Pictorial element of destination in image formation," Annals of Tourism Research, vol. 24, pp. 537-565, 1997.

[5] E. A. Frew and R. N. Shaw, "The relationship between personality, gender, and tourism behavior," Tourism Management, vol. 20, pp. 193-202, 1999.

[6] H. Xie and J. Bao, "A study of gender differences in tourist behavior," Tourism Tribune, vol. 21, pp. 44-49, 2006.

[7] D.-Y. Kim, X. Y. Lehto, and A. M. Morrison, "Gender differences in online travel information search: Implications for marketing communications on the internet," Tourism Management, vol. 28, pp. 423-433, 2007.

[8] A. M. Abubakar, M. Ilkan, R. Meshall Al-Tal, and K. K. Eluwole, "eWOM, revisit intention, destination trust and gender," Journal of Hospitality and Tourism Management, vol. 31, pp. 220-227, 2017.

[9] W. Elias, J. Benjamin, and Y. Shiftan, "Gender differences in activity and travel behavior in the Arab world," Transport Policy, vol. 44, pp. 19-27, 2015.

[10] N. López-Mosquera, "Gender differences, theory of planned behavior and willingness to pay," Journal of Environmental Psychology, vol. 45, pp. 165-175, 2016.

[11] C. Figueroa-Domecq, A. Pritchard, M. Segovia-Pérez, N. Morgan, and T. Villacé-Molinero, "Tourism gender research: A critical accounting," Annals of Tourism Research, vol. 52, pp. 87-103, 2015.

[12] C. Wang, H. Qu, and M. K. Hsu, "Toward an integrated model of tourist expectation formation and gender difference," Tourism Management, vol. 54, pp. 58-71, 2016.

[13] M. Shackley, "Tourism: A gender analysis: Vivian Kinnaird and Derek Hall (eds) Wiley, Chichester (1994) 218 pp, £39.95," Tourism Management, vol. 16, pp. 248-249, 1995/05/01/ 1995.

[14] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," Neural Information Processing Systems, pp. 993-1022, 2002.

[15] L. Azzopardi, M. Girolami, and K. v. Risjbergen, "Investigating the relationship between language model perplexity and IR precision-recall measures," presented at the Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, Toronto, Canada, 2003.

[16] B. Wu, "A study on destination choice behavior of Chinese urban residents," Acta Geographica Sinica, vol. 52, pp. 97-102, 1997.