

Research on the development of Changbai mountain summer tourism from the perspective of consumer behavior

Yanan Jin^{1,*}, and Xiaomei Li²

¹School of Economics and Management, Beijing Jiaotong University, Beijing, China

²School of Economics and Management, Beijing Information Science and Technology University, Beijing, China

Abstract: In the context of global warming, the development of the summer tourism industry has maintained a rising trend, and has gradually become a key area of widespread concern and continuous innovation by the government, tourism companies, scientific research, and the media. This paper takes domestic tourists who go to Changbai Mountain for summer travel as the research object, using the online travel notes of many travel websites such as Mafengwo and Ctrip as the data source, with the help of NLPIR big data semantic intelligent analysis platform, using text mining, network analysis, co-occurrence analysis and other methods to analyze the text data of travel notes quantitatively and qualitatively. The characteristics and laws of Changbai Mountain summer tourism consumer behavior are analyzed from three aspects: tourism decision-making behavior, tourism consumer preference, and post-tour evaluation.

1 Introduction

Under the background of the new era of mass tourism and global warming, summer tourism has ushered in a new era. Internal and external scholars gradually use UGC data to carry out tourism research, making the research results more objective and practical. Therefore, UGC data such as online travel notes have become the main research tools for more and more scholars at home and abroad, as well as an important data source for tourism consumer behavior research. This article takes Changbai Mountain tourist area as an example, takes its summer tourist consumer behavior as the research object, uses text mining and other methods to dig into the behavior characteristics and laws of Changbai Mountain summer tourist from three aspects: decision-making behavior, consumer preference and post-tour evaluation. It provides a new method and a new framework for tourism consumer behavior research, and provides decision-making reference for local government tourism planning. It has important theoretical value and practical significance.

2 Literature Review

Throughout the research on tourism consumer behavior at home and abroad, it is more inclined to conduct research on tourist motives, tourism decision-making, and spatial behavior from the individual and micro perspectives. In terms of tourism motivation, the most popular theoretical models are: distribution center model, push-pull theory, optimal arousal theory, leisure incentive method, among which push-pull theory is the most widely used [1]. In push-pull theory, the "push" factor is the internal

psychological driving factor of individual tourists' behavior [2]; the "pull" factor refers to the external driving factor of individual tourists who decide to choose one destination instead of another [3]. In terms of tourism consumer preferences, domestic and foreign research focuses on destination selection preferences (such as Tzu-Kuang Hsu), shopping preferences (such as Samuel), accommodation preferences (such as Isabel), food preferences (such as Richard), etc. [4-5]. However, they all overlooked a problem: the tourist is an organic whole, but the relevant research literature seldom studies the pre-travel, during and after traveling as a whole. This article combines the characteristics of online travel notes and related content to analyze the behavior of tourists in terms of tourism decision-making, consumer preferences, and tourist satisfaction. Different from previous studies, this article comprehensively considers the use of word frequency and keyword weights to measure tourists' concerns, uses co-occurrence analysis to dig deeper into tourists' behavior characteristics and preferences, and uses network analysis to analyze the associations between scenic spots degree, the use of sentiment analysis to measure the satisfaction of tourists, so as to obtain a deeper level of behavior data mining and analysis.

3 Tourism data collection and mining

3.1 The Development Status of Changbai Mountain Summer Tourism

The study area in this paper includes Changbai Mountain Nature Reserve and its surrounding county and township administrative areas, collectively referred to as Changbai

* Corresponding author: 19113003@bjtu.edu.cn

Mountain Tourist Area. In the summer of 2015-2018 (June-September), the Changbai Mountain tourist area received an average of 1.686 million tourists, accounting for 74.3% of the annual tourist reception. July-August is the peak tourist season, with an average tourist reception of 1.238 million tourists, accounting for about 54.6% of the annual tourist reception.

3.2 Collection and mining of tourism data

3.2.1 Sample selection of online travel data

This article is based on travel notes related to the actual tour itinerary shared online by tourists after the trip. By searching for websites with tourism community functions, and according to the research object and purpose of this article, Ctrip, Mafengwo, Baidu Travel, and Qunar.com are selected as the main data sources. This study mainly selects the online travel notes that travel from June 1 to September 30 and take summer travel as the main travel motivation. In the crawling of the text data of Internet travel notes, web crawler software (version 7.2 of the train browser) is mainly used to crawl the web travel notes data related to Changbai Mountain summer travel from the selected four online travel websites to form the original Internet travel notes database, namely Database.mdb. Then, according to the crawled travel notes database, 563 online travel notes were selected (including 254 in Mafengwo, 164 in Qunar.com, 90 in Baidu Travel, and 55 in Ctrip), which were used as the initial text database of this research. Statistics details are shown in Table 1.

Table 1. Descriptive statistics of the initial text data set

data attributes	data description
total number of travel notes	563
the travel notes span time	From 2010 to 2018, June to September each year
total words of travel notes	1156410
average length of travel notes	2050

This article summarizes and stores the 563 crawled and filtered travel notes data in an Excel table, and unifies the names of the collected travel notes data indicators. As shown in Table 2, the collected indicators include URL, travel title, destination, publication date, and departure date, travel days, cost per capita, travel companions, travel routes, travel notes, etc.

Table 2. Internet travel data collection indicators

crawl indicators	Mafengwo	Ctrip	Qunar.com	Baidu Travel
URL	√	√	√	√
travel title	√	√	√	√
travel writer	√	√	√	√
origin of trip	√	—	√	√
destination	√	√	√	√
published date	√	√	√	√
departure date	√	√	√	√

travel days	√	√	√	√
cost per capita	√	√	√	√
travel companions	√	√	√	—
travel routes	√	√	√	√
travel notes	√	√	√	√

3.2.2 Selection of text mining tools

Chinese word segmentation is the basis of text information mining. With the in-depth research of Chinese word segmentation technology, how to choose among many Chinese word segmentation systems has become a problem. By comparing common Chinese word segmentation engines, this paper selects the NLPIR big data semantic intelligent analysis platform developed by Zhang Huaping for text mining analysis.

3.2.3 Preprocessing of travel notes text data set

Due to the differences in individual characteristics of tourists and the randomness of the structure of travel notes, travel notes are expressed in different ways. In order to improve the reliability of the original corpus and the accuracy of the analysis results, it is necessary to preprocess and analyze the travel notes text data. In order to ensure the accuracy of word segmentation and word frequency statistics, the author has preprocessed the travel notes text data as follows: (1) discover new words and establish a user dictionary; (2) establish a stop vocabulary list and clean up the data; (3) circulate word segmentation and word frequency statistics. Finally, the results of word frequency statistics on the characteristics of Changbai Mountain summer tourism consumer behavior are sorted out. After the correlation analysis is eliminated, some of the results are shown in Table 3.

Table 3. High-frequency word statistics of the text data of Changbai Mountain's summer travel notes (partial result)

High-frequency words	weights	word frequency	unary probability	information entropy
Changbai Mountain	724.61	4685	0.014263	0.060619
Tianchi	455.01	3259	0.010257	0.046974
scenic spot	313.97	2025	0.006223	0.031612
hotel	352.07	1787	0.005935	0.030428
time	283.09	1556	0.003831	0.021317
xipo	263.07	1438	0.004973	0.026373
beipo	279.53	1317	0.004653	0.024988

4 Analysis of Tourism Consumer Behavior Based on UGC Data

4.1 Analysis of tourist decision-making behavior

Travel decision-making is the external manifestation of

travel motivation and the core behavior in the travel process. According to the analysis of travel notes and word frequency, the characteristics of travel decision-making behavior are mainly analyzed from three aspects: travel decision-making behavior, decision-making information demand and information acquisition channels.

This paper extracts the verbs related to decision-making behaviors, as shown in Table 4. After content analysis, it is found that tourists will go through the process of query/search-selection/decision-planning-reservation before the tour. Before traveling, tourists need to make decisions about choosing a tourist destination, planning a travel route or itinerary, booking tickets and accommodation. Therefore, tourists need travel information about tourist destinations or scenic spots, tourist routes or itineraries, transportation, accommodation, weather, etc. before traveling. After keyword analysis and co-occurrence analysis (see Tables 4 and 5), the main sources of tourism information acquisition are the Internet, including travel notes or APPs of travel websites such as Ctrip, search engines such as Baidu, and third-party platforms such as Taobao and Dianping. And for different information needs, tourists have different preferences for choosing information channels. For example, tourists mainly obtain information on destinations and scenic spots, travel routes, itineraries, etc. through the guides, travel notes, and reviews of travel community websites; tourists book and purchase tickets through the scenic spot official website or APP, online travel websites, Taobao, etc.; tourists book accommodation through the official website of the scenic spot or Ctrip.com; tourists use Dianping and other platforms to inquire about recommended foods, etc.

Tables 4. Keyword extraction and analysis of travel decision-making behavior

keywords	weights	word frequency
selection	166.59	710
plan	105.26	269
decision	91.35	290
query/search	30.12	109
reservation	27.13	395

Tables 5. Keyword extraction and analysis of tourism information channels

keywords	weights	word frequency
strategy	103.56	251
Changbai Mountain Yitong APP	57.99	77
ctrip.com	50.91	92
qunar.com	42.73	92
dianping	27.48	34
internet	22.37	371

4.2 Analysis of Tourists' Consumer preference

4.2.1 tourism traffic

Due to the high cost of airfare during the peak season and the distance, most tourists choose railways as long-distance transportation. Among them, more comfortable sleepers and fast high-speed railways have become the choice of most people. After considering the route, distance, fare, time and other factors, some tourists choose the right airline. If there is no suitable direct flight, they will reach their destination by transferring or "airplane + train" combined transportation.

4.2.2 tourism catering

According to the co-occurrence analysis, the dishes of stewed iron pot (60), stewed vegetables (54), stewed pork in pot (35) and cold-water fish (24) are mentioned more frequently and are relatively popular among tourists. In tourists' impressions of Northeastern cuisine, the words "characteristic", "large portion", "hostile", "authentic" and "rough" have been mentioned many times. The second is North Korean/Korean-style dishes, including cold noodles (154), barbecue (103), dog meat (77), spicy cabbage/kimchi (59), bibimbap (44), walleye pollack (43), and rice cakes (43), rice wine (41) is a special dish that tourists taste more. The authentic foreign flavor is the impression of tourists on North Korean cuisine. In addition, hot spring eggs (264) and hot spring corn (106), the specialty snacks in Changbai Mountain Scenic Area, have become one of the must-see delicacies for tourists visiting the North Slope due to their high popularity and unique taste.

4.2.3 tourism shopping

Travel shopping behavior is one of the main behaviors of tourists in the travel process. Changbai Mountain is well-known both at home and abroad for its rich variety of local specialties, thus, special products (65) and gifts/souvenirs (32) are the main purchases of Changbai Mountain summer tourists. Among the specialty products of Changbai Mountain, blueberries (126), ginseng (80), mushrooms (52), and fungus (28) are the specialty products that tourists prefer to buy. Among them, blueberries and ginseng are tourists' favorite tourist products, and Changbai Mountain postcards (21) are the most popular souvenir for tourists.

4.3 Analysis of tourist evaluation behavior

Through the emotional score statistics, the average emotional score of the 563 travel notes is 46.73 points, of which the positive sentiment average score is 87.04 points, and the negative sentiment average score is -40.31 points. As shown in Fig.1, according to the emotional score to measure different emotional intensity: positive emotion (5, +∞), negative emotion (-∞, -5) and neutral emotion [-5, 5], of which positive emotion can be divided into high (25, +∞), moderate (15, 25] and general (5, 15], negative

emotions can be divided into high ($-\infty, -25$), moderate $[-25, -15]$ and general $[-15, -5]$.

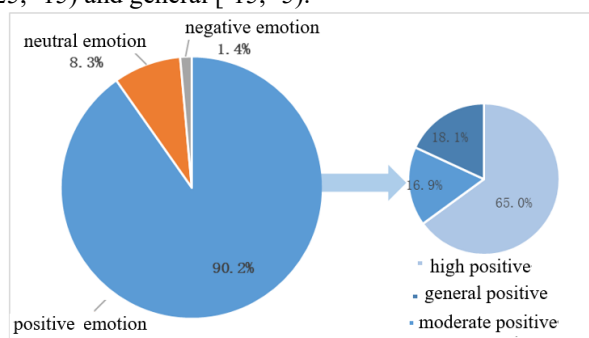


Fig.1. Statistical chart of emotional tendency analysis of travel notes

According to the statistical results in Figure 1, 90.2% of tourists have positive emotional tendency towards the summer travel experience in Changbai Mountain, among which 65% have the highest degree of positive emotional tendency, 16.9% are moderate and 18.1% are general. Only 1.4% of the tourists' emotional tendency is negative, and all of them are highly negative, which indicates that most of the tourists' emotional tendency towards the summer travel experience in Changbai Mountain is

positive.

5 conclusion

The research on tourism consumer behavior based on UGC data in this paper can help scenic spots and other tourism enterprises to understand the consumer preferences of different tourist groups. At the same time, tourism websites can also be helped to establish a personalized scenic spot recommendation system according to the consumer preferences of tourists.

References

1. M. Caber, T. Albayrak. *TOURISM MANAGE*, **55** (2016).
2. M.Y. Wu, P.L. Pearce. *TOURISM MANAGE*, **43** (2014).
3. D.B. Klenosky. *J TRAVEL RES*, **40**, 4 (2002).
4. T.K. Hsu, Yi. F. Tsai, H.H. Wu. *TOURISM MANAGE*, **30**,2 (2009).
5. S. S. Kim, D. J. Timothy, J. Hwang. *TOURISM MANAGE*, **32**,3 (2011).