Portrait positioning and demand analysis of bridge fusion target population based on internet big data

Jiawei Gao*[a], Yinan Cheng[a], Jie Liu[a], Chao Chen[b]
[a]Planning and Research Institute of Ministry of Transport, Beijing 100028, China; [b]Guizhou Lu'an Expressway Co., Ltd., Lu'an 562200, Guizhou, China

ABSTRACT

In order to explore and enhance the accuracy and precision of target group positioning for highway bridge tourism, as well as to provide a more personalized and diverse experience of bridge tourism integration, this article will focus on the lack of research and practice in crowd portrait methods within the field of bridge tourism integration. It aims to utilize Internet big data resources, relying on technologies such as data collection, information extraction, classified statistics, and portrait generation to analyze the characteristics of the target population from two dimensions: tourists' basic needs and expansion needs. Once the characteristics of the target population are defined, their travel demand, consumption demand, and tourism demand will be further analyzed. This analysis will lay a foundation for future functional positioning and format design of bridge tourism projects.

Keywords: Bridge tourism, portrait of target group, internet big data, functional positioning, format design

1. INTRODUCTION

Crowd profiling involves categorizing user information based on various attributes such as behavior patterns, personal preferences, and other data points to provide a unified description that captures valuable statistical characteristics. This approach allows for summarizing useful information about user groups based on their attributes. Currently, crowd portrait technology is primarily practiced in the fields of electronic information, finance, business, publications, leisure travel and so on[1]. Some studies conduct tourist crowd portraits through questionnaires, tourist evaluations, tourism sharing and other data[2]. In summary, at this stage, the target population portrait method has been widely used in the preliminary market analysis of the tourism industry and has been effectively combined with advanced intelligent means such as big data, Internet and AI[3]. However there is still a gap in the field of bridge and tourism integration where it fails to effectively guide the deep integration of bridge and tourism industries from a data level perspective. It is one of the important methods to improve accuracy scientificity and wisdom of bridge tourism integration in future.

Based on the background of bridge tourism and the development characteristics of crowd portrait methods, this study utilizes deep extraction of Internet big data to conduct statistical analysis and mining from various aspects such as personal basic information, travel demand information, tourism purpose information, and consumption preference information. The study outlines the results of individual portraits of bridge tourist groups in order to identify their travel demand, consumption demand, and tourism demand. Using the Huajiang Canyon Bridge of Lu'an Expressway in Guizhou Province as a case study, portrait and demand analysis were conducted on the future target population of the project. This provides support for improving the functional positioning and business mode selection of bridge engineering.

2. CHARACTERISTICS AND REQUIREMENTS OF BRIDGE TOURISM TARGET POPULATION PORTRAIT DATA

Data collection is an important basis for the analysis of bridge travel integrated crowd portrait. The real-time and accuracy of data information directly determine the practicability, applicability and effectiveness of travel information service judgment.

*2351998603@qq.com; phone 18610005629; fax 010-57802970

International Conference on Optics, Electronics, and Communication Engineering (OECE 2024), edited by Yang Yue, Proc. of SPIE Vol. 13395, 1339540 · © 2024 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.3049771

2.1 Determination of the data requirements

In order to systemize and order all kinds of information resources and facilitate sharing, retrieval and expansion, the data resources required by each application system are classified, merged and fused, and the data to be collected for bridge and travel integration is determined to include basic data and business data[4].

2.2 Data collection scheme

(1) Data interconnection. Formulation of the specification of data interface; Docking data interface development of various units, including but not limited to traffic and tourism industry management department data interface, road passenger/terminal passenger information interface, railway passenger information interface, air passenger information interface, etc.

(2) Data collection, cleaning and desensitization. Construction of Web crawler system, crawling traffic and tourism website data, and real-time upload cloud platform for processing; Development of data extraction, data filtering, data conversion, data desensitization, data supplement and other modules to realize the collection and cleaning of configurable and automated operation.

(3) Data management. Development and establishment of process management, version management, scheduling management and task management a total of 4 data management function modules (Refer to Figure 1 for further information).

(4) Data quality monitoring. Performance of 4 kinds of data inspection, including data entry verification, key data audit, data entity and correlation inspection, key indicator inspection; Establishment of data quality evaluation index system, including correction for missing values, inspection for outliers, and visualization for monitoring.

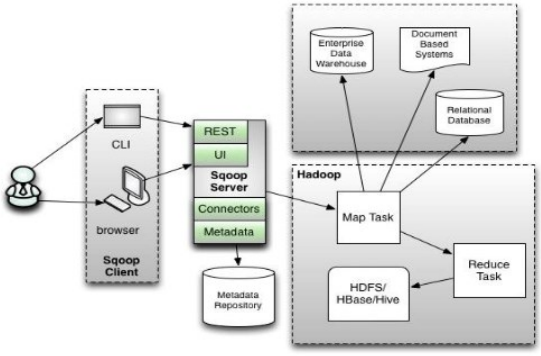

Figure 1. Schematic diagram of data acquisition process for Internet implementation.

3. BRIDGE TOURISM TARGET POPULATION PORTRAIT METHOD

3.1 Data processing

The core work of data processing is to load the data collected from the Internet into the buffer[5]. The data buffer can be in file or database mode; If database is used, temporary data and official data are stored separately. Data filtering is responsible for eliminating “dirty data” and eliminating data inconsistencies. “Dirty data” includes irregular data, data that does not conform to the facts, such as: value range, integrity rules, spell check, etc. Missing data can be processed by global constants, attribute means, possible values, or simply ignoring the data. Noise data can be removed by methods such as box division (grouping the original data, and then smoothing the data within each group), clustering, computer manual inspection and regression. For inconsistent data, manual correction can be performed. The content of data should be encrypted or desensitized, and the role and ownership of system management should be separated; The data is encrypted to ensure that it cannot be viewed by the comprehensive system administrator, while the data is encrypted with the token of the data owner, and the data content is technically impossible to snoop on without the authorization of the data owner. Therefore, the technical administrator who maintains the availability of the system does not have the ability to access the data content, and the role of system administration is well separated from ownership.

3.2 Prediction of bridge tourist population size

In the prediction of the bridge tourist population, the occurrence and attraction of bridge tourism are predicted in addition to the total number of bridge tourism based on the statistics of population, social and economic data and tourism survey data. At present, under the premise of stable social and economic development, and no major policy changes, the local tourism market continues to recover, consumer confidence is effectively boosted, and consumer demand is accelerated. Therefore, this study adopts the elastic coefficient method to forecast the total number of bridge tourists.

$$VF = (AT/(AB + AF)) \times (VT + VB) \quad (1)$$

VT: Annual visitor count for similar tourist attractions

VB: Annual visitor count for similar bridges

VF: Predicted annual visitor count for the target bridge tourism project

AT: Attraction coefficient of similar tourist attractions

AB: Attraction coefficient of similar bridges

AF: Attraction coefficient of the target bridge tourism project

4. DEMAND MODEL OF BRIDGE TOURISM CROWD

Based on the experience and case analysis of existing bridge tourism projects, it can be seen that tourists' needs are mostly divided into the following seven categories[6].

4.1 Theme block mode

This mode is to create a theme block at the node, with the help of IP, restore scenes from movies, TV, novels, legends, folklore and other cultural resources, and set up dramatic dining, accommodation and shopping to simulate the sense of time travel, attract tourists to enter the project and experience a unique lifestyle, in order to confirm tourists' imagination of cultural IP, and provide unique services with "surprise" and "move". Through strong visual impact, storytelling, cultural and creative experience formats, stimulate consumer feelings, strengthen impression, extend consumer stay time, improve consumption, and realize brand communication.

4.2 Characteristic town model

The characteristic town mode is to upgrade the traffic node from the traffic service function to the leisure tourism consumption function[7]. Expand to "leisure, entertainment, shopping, dining, accommodation", so that tourists can stay longer at the traffic node, have more ways of leisure, have more choices of shopping, and have more combinations of tourism projects, so that they can relax both physically and mentally. The town needs to concentrate local (county) specialties, famous snacks and packaging county culture, such as: minority culture, unique agricultural products and agricultural and sideline products, unique crafts and handicrafts, etc., so that the town has a unique and non-replicable. Strengthen the characteristics, and form a consumer experience support system. At the same time, create a characteristic town brand, and aim to become the first or only choice to lock the theme consumption experience.

4.3 Shopping and leisure mode of characteristic industry

The bridge node displays the local characteristic industries from all angles and launches a series of characteristic products to attract tourists to stay and consume; Gather famous boutiques, specialty stores and specialty stores to form characteristic boutique shopping blocks. Gather the top ranked famous shops in the surrounding area and related areas, as well as the local unique products, such as handicraft products, agricultural and sideline products, agricultural products, etc., to form a good shopping environment, so that travelers at the bridge node can buy local famous brands and specialties here, without having to go down the road to purchase.

4.4 Commercial complex model

Commercial complex is a comprehensive commercial collection (also known as shopping mall) that provides customers with one-stop consumption of shopping, entertainment, leisure, catering, etc. It is suitable for the bridge node of the high-flow highway in the suburbs of the city or between two major central cities. It has convenient transportation and is easy to gather industries and population, which makes the land have commercial development potential. According to the specific

characteristics of the land and location advantages, the commercial complex is built to become a part of the city's commerce.

4.5 Industrial base model

Industrial bases with different theme positioning such as cultural and sports industrial park, film and television industry base, and cultural and creative base can be built according to the situation and needs of industrial resources around the bridge node.

4.6 Logistics service model

Bridge nodes not only have the flow of people and logistics, but also business flow, information flow, financial flow and cultural flow. Therefore, make full use of the advantages of expressways to carry out comprehensive logistics services such as warehousing, distribution, transit and packaging.

4.7 RV campsite

In order to meet the growing demand of self-driving tourists, relying on the high-quality landscape resources around the bridge node[8], the car/RV camp is built to integrate ecological tourism, camping, shopping, expansion, bonfire, barbecue and other leisure projects, and has a comprehensive reception service area, landscape leisure area, RV camping area and outdoor expansion area to meet the needs of different tourists.

5. CASE STUDY

5.1 Project overview

Huajiang Canyon Bridge is located at K61~K64 section of Lu'an Expressway in Guizhou Province, across the Huajiang Grand Canyon. As a demonstration project of bridge and tourist integration in Guizhou Province, Huajiang Canyon Bridge will become a "bridge and tourist integration" multi-functional complex integrating bridge sightseeing and high-altitude adventure, driving the development of local tourism economy.

5.2 Data collection and screening of target population

This study focuses on building a basic data base (Refer to Table 1) based on cross-reference data, mobile signaling data, user-generated content (UGC) statistics from online social media platforms and tourism data collected from statistics bureau, tourism Bureau and other departments.

Table 1. Platform for extracting internet data and its requirements.

Web Search Data	The data are searched on Baidu, 360, Sogou and other platforms to analyze tourists' motivation and popularity on the Internet
Operator data	Mobile signaling data, which can analyze the specific spatial and temporal composition of tourists and the number of visits and other refined data dimensions. According to the mobile signaling data, the mobile signaling data can analyze the entry mode and visit situation of tourists in the region
Unionpay data	Analysis of UnionPay consumption data to assess the popularity and spending capacity of tourists.
OTA platform data	UGC data of major OTAs such as Ctrip, Hornet's Nest and Meituan, analyzing tourist perception and tourist attributes
Other data	Dadi Yunyouwen Travel database, LBS positioning data, microblog Tiktok and other we-media data are analyzed to analyze the popularity of tourists' visits and punch card propagation

(1) Visitor source index

Based on the customer source data of Sky Bridge service area and Anshun Balinghe Bridge, we conduct behavioral analysis of Internet search popularity and search concerns. And then we integrate search engine data, LBS positioning data, OTA platform data, tourist transaction data and other multiple data to conduct a comprehensive analysis of the project tourist market, such as tourist source market, market leakage, search popularity, search concerns, etc., to fully understand the market trend and preference of project tourists.

Using the data from the Balinghe Bridge and Sky Bridge in 2023 as a reference, it is forecasted that the Huajiang Canyon Bridge will host an estimated 12.906 million visitors in its first year of operation.

(2) Tourist portrait

The first step is to comprehensively analyze tourists based on tourism big data, including tourist attributes and behavior analysis. The analysis dimension includes tourists' gender, age, occupation, consumption level, interest preference, travel mode, taste preference, search focus, associated travel destination, destination type preference, travel motivation, partner mode, travel time, travel days with group tour, etc., The analysis result can accurately locate the market demand of tourists, and propose the main customer groups and product/business development direction of the project in the future.

Based on the supply data of business forms around the Sky Bridge service area and Anshun Balinghe Bridge, the supply and distribution of tourism supporting business forms such as scenic spots, hotels and catering were analyzed. The popularity analysis results are shown in Table 2.

Table 2. The search popularity of the Balinghe Bridge, Sky Bridge, and Huajiang Canyon Bridge in 2023 on the internet.

Keyword	Overall daily average	Overall year-on-year
Balinghe Bridge	236	17%
Sky Bridge	7378	376.82%
Huajiang Canyon Bridge	502	96%

(3) Market perception

With tourism UGC data, Douyin Weibo and other self-media data as the main data sources, LBS positioning data was integrated to carry out combined data analysis with Sky Bridge service area and Anshun Balinghe Bridge as the core and surrounding key scenic spots as attractions. The dimensions include word-of-mouth popularity, word-of-mouth public opinion, negative labels, etc. Based on the 5A-level scenic spot evaluation standard and combined with tourist semantics, negative semantic labels are carried out on the Sky Bridge service area, Anshun Balinghe Bridge and the surrounding key scenic spots to analyze the gap. The advantages of the project in cultural tourism market and potential power are proposed by means of data, and relevant development suggestions are also given. The analysis results of tourist markets are shown in Table 3.

Table 3. Balinghe Bridge, Sky Bridge top 10 tourist markets, Huajiang Canyon Bridge top 10 potential tourist markets.

Balinghe bridge		Sky bridge		Huajiang Banyon bridge	
Urban	Standardized Heat Index (USHI)	Urban	Standardized Heat Index (USHI)	Urban	Standardized Heat Index (USHI)
Guiyang	98.00	Guiyang	98.00	Guiyang	98.00
Qiandongnan Miao and Dong Autonomous Prefecture	35.08	Qiannan Buyi and Miao Autonomous Prefecture	26.42	Anshun	45.67
Anshun	17.15	Chengdu	16.51	Qiandongnan Miao and Dong Autonomous Prefecture	23.10
Chongqing	15.37	Liupanshui	15.86	Chongqing	11.60
Qiannan Buyi and Miao Autonomous Prefecture	8.22	Qianxinan Buyei and Miao Autonomous Prefecture	13.46	Shanghai	11.95

Balinghe bridge		Sky bridge		Huajiang Banyon bridge	
Liupanshui	8.04	Qiandongnan Miao and Dong Autonomous Prefecture	11.13	Chengdu	7.72
Chengdu	7.13	Chongqing	10.81	Qianxinan Buyei and Miao Autonomous Prefecture	10.84
Guangzhou	6.73	Guangzhou	7.82	Beijing	5.17
Bijie	5.55	Beijing	5.40	Qiannan Buyi and Miao Autonomous Prefecture	14.20
Tongren	4.13	Kunming	4.95	Guangzhou	4.30

5.3 Accurate portrait results of project population

In this study, Huajiang Canyon Bridge was taken as the research object, and a total of 5426 pieces of relevant data and 8453 pieces of comment information were collected from the travel OTA platform by using Internet online crawler technology. After the relevant data and comments were fully screened by computer algorithm, 2634 pieces of relevant data and 5436 pieces of comments were retained. Then, the useful information is extracted and classified according to the words about travel, tourism and consumption in the user evaluation text, emotional expression and attributes in the travel photos.

Based on the above data analysis, the characteristics of the crowd portraits of Huajiang Canyon Bridge are as follows: According to the tourism service conditions and attractiveness of the region where the project is located, the future tourists of this project will mainly come from the province, especially the surrounding Guiyang, Anshun and Xingyi. Due to the lack of tourism brand awareness in the region and the poor conditions of accommodation and other supporting facilities, the current tourist source area has a small radiation radius and is only a regional tourist destination. In the future, through the comprehensive development of the region, Beipanjiang Grand Canyon and this project, it can further improve the attraction of medium and remote domestic markets and inbound markets, and expand the coverage radius of tourist source area. Lu'an Expressway will become a large flow channel in the future, with more transit tourists. However, due to the lack of experience projects, brand awareness and lack of accommodation facilities in the area, it is expected that the tourists of this project will stay for a short time. In the next step, by strengthening the "night sightseeing" experience, this project will enhance the overall guest retention capacity of the region and promote the development of overnight tourism economy.

6. CONCLUSIONS

Currently, research on crowd portraits in China's tourism industry primarily utilizes methods such as mass surveys and statistical analysis. This study applies Internet big data analysis and relevant information crawling screening methods to analyze and research crowd portraits in bridge tourism projects. The aim is to uncover the needs of tourists and summarize their diverse experiences in bridge tourism. As a result, a method and process for constructing the target tourist portrait in bridge and tourism integration projects are innovatively developed: (1) Building a bridge tourist crowd portrait analysis method based on Internet data, while simultaneously creating a prediction model for the size of the bridge tourist crowd; (2) Analyzing the travel demands of tourists for Huajiang Canyon Bridge according to the results of crowd portraits, providing valuable insights for the construction of the bridge tourism project based on the needs of its construction unit; (3) Utilizing Internet data analysis to develop a target demand model for bridge tourism that effectively enhances the relevance, scientific rigor, and feasibility of such projects. This study also categorizes seven modes of bridge and tourism integration, comprehensively improving overall service levels within this sector. Furthermore, it scientifically supports efforts aimed at enhancing technical quality within key construction projects related to bridge tourism with an eye towards achieving significant practical impact. Ultimately, this study provides valuable technical references for subsequent practical projects.

ACKNOWLEDGMENTS

Fund Project: Science and Technology Plan Project of Guizhou Provincial Department of Transportation (2022-122-027).

REFERENCES

[1] Luo, W. and Yang, X., "User profile construction in tourism field based on travel notes data mining," Tourism Overview 1, 24-27 (2022).
[2] Wang, L., Shen, Z. and Li, Y., "User profiling of social Q&A community," Information Studies: Theory & Application 41(1), 129-134 (2018).
[3] Xu, P., Bi, Q., Zhang, H., et al., "Construction of digital library user profile driven by data," Library and Information Service 63(3), 30-37 (2019).
[4] Sun, Z., Luo, Y., Zheng, X., et al., "Intelligent travel route recommendation method integrating user emotion and similarity," Computer Science 48(S1), 226-230 (2021).
[5] Badawy, M. K., "Is open innovation a field of study or a communication barrier to theory development: A perspective," Technovation 31(1), 65-67 (2011).
[6] Xanthopoulos, I., Goulas, G., Gogos, C., et al., "Highway rest areas simultaneous energy optimization and user satisfaction," Pan-Hellenic Conference on Informatics (2016).
[7] Ettema, D., Garling, T., Olsson, L., et al., "The road to happiness: measuring Dutch car drivers' satisfaction with travel," Transport Policy 27, 171-178 (2013).
[8] Zeng, C., Lyu, B., Deng, S., et al., "Benefits of a three-day bamboo forest therapy session on the physiological responses of university students," International Journal of Environmental Research and Public Health 17(9), 3238 (2020).