Research on the Construction and Application of Smart Tourism Platforms Based on Big Data Technology

Yongjun Zhang(✉)

School of Information Technology, Key Laboratory of Intelligent Information Processing and Information Security in Universities of Shandong, Shandong Youth University of Political Science, Jingshi East Road 31699, Jinan, China
Yongjunzhang@sdyu.edu.cn

Abstract. With the advent of the mass tourism era, innovative tourism services have attracted increasing attention from numerous tourism consumers. The wide application of big data and cloud computing technology makes it possible to build an intelligent tourism service platform to improve the experience of tourists. This research proposes a series of building technical proposals for a smart tourism service platform based on big data technology. The platform mainly focuses on massive data-based tourism administration, such as emotion analysis, tourist route recommendation, tourist portrait, passenger flow trend prediction, using big data analysis, mining, and processing. It also produces a series of relevant services and constructs special applications with various functions to realize tourism data acquisitions, research, and manipulations. Finally, the intelligent tourism platform's massive data analysis and processing results provide feasible and optimized decision-making references for the government, enterprises, tourists, and other stakeholders. It puts forward countermeasures for the innovative tourism service platform's implementation and application to promote the sustainable development of creative tourism.

Keywords: Big data · Smart tourism · Tourism platform

1 Introduction

In the era of big data, it has always been a topic of great concern in the industry to fully discover the new changes and trends in contemporary tourism development. With the rapid growth of tourism and the acceleration of tourism informatization, the traditional tourism system has been unable to satisfy the overall requirements of the rapid development of the modern tourism industry (H Li&T Su,2020). Meanwhile, the traditional tourism data analysis and management systems gradually become incompetent for high-density tourism business processing with dramatic-produced massive data from tourism business. They do not sufficiently provide high-quality service and reasonable decision-making assistance for tourists (Chang Xu, 2018). Therefore, extensive data analysis and tourism management technology provide high-quality development ideas

G. Guan et al. (Eds.): ICBDSS 2022, AHCS 8, pp. 562–573, 2023.
https://doi.org/10.2991/978-94-6463-064-0_58

for service-oriented tourism and offer a new direction for the intelligent tourism industry to increase tourists' personalized experience.

Smart tourism aims to seek better design and systematic management, effective methodology, and adopt effective resource management methods in integrating new value propositions for tourists, tourism, and the hotel industry (Y.Wu, 2021). The diversified technologies, such as mobile applications, sensors, data analysis, optimization, and visualization, have become a reality based on the innovations of smart tourism. It contended that tourism demand forecasting is critical in the application and dramatically influences tourism practitioners (Hengyun Li, 2020). The planning, operation, and management of tourist attractions heavily depend on tourist trends predicting (Huang, Zhang, & Ding, 2017). Some studies are to make the big data management system more perfect by developing and utilizing the 5G Network and Internet of Things (Haifeng Gao, 2021). In addition, some researchers proposed a recommendation system based on big data technologies that can realize the tourist attractions and be tailored to tourist preferences for promoting the local tourism industry (Khalid Al Fararni, 2021).

Massive data analysis and process in tourism applications can provide valuable data for tourism departments, tourism-related enterprises, and tourists. The collection, research, and application of big data can promote the transformation of new and old driving forces of tourism and greatly promote the development of the tourism industry.

In conclusion, smart tourism and its big data will promote, develop and integrate each other with the rapid innovation and integrated development of the tourism industry. For one thing, smart tourism encourages the generation of big tourism data, which includes massive data of scenic spots, hotels, travel agencies, tour guides, tourists, and tourism enterprises. For another, the application of tourism massive data analysis promotes the intelligent process of tourism, such as tourism public opinion monitoring and analysis, tourism passenger flow monitoring and analysis, tourism consumption monitoring and analysis, and OTA (Online Travel Agency) consumption monitoring. Combining big data and intelligent tourism is the inevitable trend in developing the tourism industry.

2 Demand Analysis on the Construction of a Smart Tourism Platform Based on Big Data Technology

The bright tourism platform, with the help of cloud computing, big data, and portable mobile devices, arms actively collect and monitor the tourism-related resources and activities and publish that information in time to let people know relevant information to arrange and adjust work and tourism plans in time by effectively utilizing tourism information. The development of intelligent tourism platforms continuously generates a series of dynamic massive tourism data, such as tourism website data, tourism mobile app data, tourism new media data, tourism video data, virtual tourism data, etc. These critical and valuable data should be collected, stored, processed, and applied for further analysis. In addition, multi-modal data fusion can provide robust data support for model decision-making and improve decision-making accuracy. It integrates the advantages of each input data set and makes up for the shortcomings of a single data set by using the complementarity of various data sets in terms of space-time resolution, integrity, and accuracy through a particular mathematical algorithm. Compared with single-source

Table 1. The multi-source data fusion method in tourism data acquiring and processing.

Data type	Fusion method	Fusion algorithm
Multi-source data	Data layer fusion, Feature layer fusion, Decision level fusion	①Component substitution (CS), multi-resolution analysis (MRA), model-based algorithm; ②Bayesian estimation, D-S evidence theory, clustering algorithm, neural network algorithm; ③Identification based decision fusion method: map method, ML method; ④Knowledge-based decision fusion methods: Expert knowledge method, Neural network method, Support Vector Machine (SVM)
Multi-modal data	Fusion Models	①Multi kernel learning (MKL) method: Object Classification, Emotion Recognition; ②Image model (GM) method: Dual-mode speech, Emotion Recognition, media classification; ③Neural network (NN) method: Emotion Recognition, Dual-mode speech

processing alone, data fusion has higher detectability and reliability, a more comprehensive space-time perception range, lower reasoning ambiguity, more target features, and more robust fault tolerance. Multi-modal data fusion aims to establish a model to process and correlate information from multiple modes (Sen Qiu,2021). Table1 lists the multi-source and multi-mode data fusion methods.

3 Construction of Smart Tourism Platform Based on Big Data Technology

The smart tourism service platform based on big data is to build an intelligent tourism application based on multi-dimensional and massive business data. It is a platform for tourism data sharing, exchange, and application. Based on the collection of tourism information, it synthetically integrates tourism industry information. It also profoundly carries out the analysis and application of big tourism data to realize intelligent management, marketing, and service of the tourism industry. Establishing the innovative tourism platforms generally adopts big data technology, builds four layers of services (Basic service layer, Data service layer, Platform service layer, Software service layer), and makes multiple thematic applications. Each thematic application closely depends on big data and performs its duties. It can integrate scenic spots, accommodation, transport, catering, entertainment, commerce, leisure, and other related recreational resources, collect, store, clean, search, exchange, manage and classify the tourism data. In addition, the data analysis and processing results also provide decision-making references for the government, tourism enterprises, tourists, and other stakeholders. The platform can analyze and apply wise tourism statistics, marketing, management, service, price monitoring, credit, etc. It dramatically improves the standardization level of tourism management.

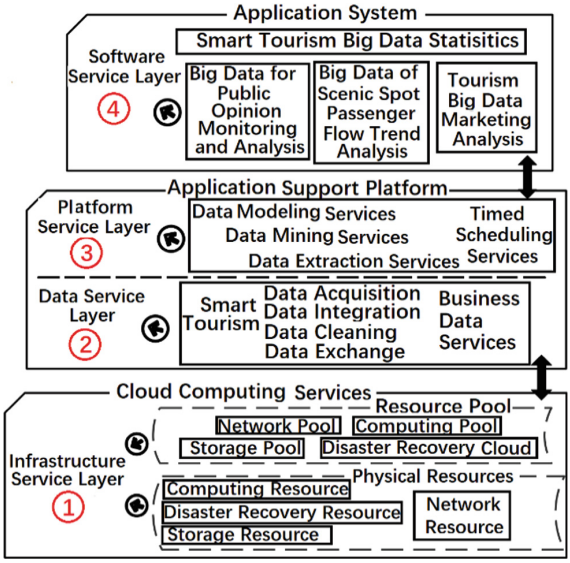

Fig. 1. Architecture diagram of big data-assisted smart tourism platform.

The platform also provides essential support for government decision-making, sharing the information of tourism enterprises, providing accurate information services for public traveling, improving the tourism industry's intelligence level, and promoting the high-quality development of the tourism industry. The architecture diagram of the big data-assisted innovative tourism platform is shown in the following Fig. 1.

3.1 Overall Architecture

3.1.1 Infrastructure Service Layer

The infrastructure service layer is the bottom layer of the platform, including essential condition service functions such as network, cloud computing, and ample data storage. It adopts virtualization solutions with simple operation and high operation and maintenance efficiency. It also has a disaster recovery pool, realizes multi-data backup, and ensures data security. It is a fundamental support platform with dynamic expansion, flexible configuration, safety, and stability. According to the installation rules of the Hadoop cloud service platform, the cluster of tourism service platforms shall be divided into NameNode and DataNode during hardware construction and integration. NameNode stores metadata for recording data fragments and managing and distributing computing tasks. DataNode keeps data fragments and performs calculation tasks distributed by NameNode. After completion, it is summarized into NameNode for consolidation and assembly. In the cloud service platform, Gigabit enterprise switches can be selected, which can control the hardware cost of the service to the maximum extent based on ensuring the operation reliability and performance of tourism management.

3.1.2 Data Service Layer

Data storage is the basis of data analysis and visualization. The data service layer has a high-performance and highly available storage server, which can realize many different types of data warehouses. It collects the business data of each system and exchanges the data by using big data technology to share the data of each design fully and finally eliminate the information island. At the same time, the data service layer can realize the collection and processing of data and improve the value of data. The distributed storage model is adopted for the data storage of tourism big data. By establishing a file server, picture server, relational database, and non-relational database, the continuous availability and rapid query of data are ensured. Standard open-source tools such as Hadoop, HDFS, HBase, etc., use these storage technologies to provide real-time data extraction and multi-point data association analysis. Tourism big data includes spatial data and text data. Text indexing technology adopts the method of inverted indexing, which is an efficient indexing method for massive text retrieval and mapping the storage location of text data from the tourism business. Spatial index technology will divide the spatial area with an orthogonal grid and map any object in the space to a specific grid cell with the corresponding mapping function. It combines the inverted index and spatial grid division to form an inverted grid index and realize the algorithm of the index structure on the MapReduce distributed model.

3.1.3 Platform Service Layer

The platform service layer mainly realizes application management and provides a unified, open, and shared cloud service platform for application development. It supports various forms of application access such as web application, mobile application, and desktop application, application creation, development, testing, audit, online operation, and maintenance. In addition, the general functional services provided by the platform mainly include a one-stop service, cloud search service, computing model, statistical service, and unified user management. The applications connected to the platform can effectively use the available services and enhance the value of applications. Intelligent computing of tourism big data is the core application of the management platform, which includes data mining and predictive analytics data mining technology. The innovative big data analysis system can be configured with two methods: streaming data calculation and memory data calculation, requiring relevant response components. Streaming computing responds to the data by receiving and popping up and starting the analysis. In the analysis process, the configuration and algorithm are first injected into the algorithm components, the storm framework is used for streaming analysis, and the calculation results are output. In the response process of memory computing, memory control shall be carried out to store the data in memory orderly. Then, combined with configuration and algorithm, the data shall be analyzed through the spark framework. Finally, the calculation results shall be output. Table 2 shows the intelligent computing technology classification of big tourism data.

Table 2. The classification of key technologies of big data intelligent computing.

Key technology	Analytical method	Typical technical analysis algorithm
Data Mining	Cluster Analysis; Classification and Forecasting; Association analysis	① Division method: K-means, K-center algorithm; Hierarchical methods: birch method, CURE method, and CHA MELEON method; Density-based methods: DBSCAN algorithm, options algorithm, DENCLIUE algorithm; Network-based method: CLIQUE algorithm, STING algorithm; model-based approach: EM algorithm; ② Decision tree, Rough Set, Bayesian, genetic algorithm, neural network algorithm; ③ Apriori algorithm, a frequent pattern tree algorithm
Forecast Analysis	Qualitative prediction; Quantitative prediction	① Boosting, Bayesian network algorithm; ARIMA, Markov, Multiple linear regression, Grey prediction ② Exponential smoothing method, trend extrapolation method, and moving average method; Linear regression; Machine learning algorithm: LSTM, SVM, PSO

3.1.4 Software Service Layer

The software service layer is mainly oriented to the unique tourism businesses to realize the relevant transactions, such as holiday passenger flow analysis and prediction, tourist value analysis, passenger flow monitoring analysis, satisfaction index analysis, etc. The system can operate independently and realize data exchange and sharing and constantly develop the functions of the intelligent tourism platform. Meanwhile, the software application can be dynamically expanded and updated over time.

Data visualization is the presentation of data. Through the interactive visual interface, the results of data analysis are transparent and materialized. Big data visualization allows users to visually see the tourism data study results, retrieve the items they need and solve their problems. It provides data support for scheme decision-making and makes scientific predictions for the trend of the situation. ChronoViz, D3, FlightGear, and Highcharts are the commonly-used intelligent visualization software to facilitate massive tourism data analysis and management. Figure 2 illustrates the data visualization analysis of the tourism big data management platform.

Fig. 2. Data visualization analysis of tourism big data management platform.

3.2 Security Architecture

Security architecture is one of the technical forces which mainly ensures data security and prevents data loss, error, attack, and other problems. It primarily deals with data security threats from three aspects: storage, application, and management of tourism big data. Figure 3 illustrates the security architecture for the proposed intelligent tourism platform in detail.

- Data storage layer. Data storage is the essential condition layer, which mainly realizes data encryption functions, vital separation, filter use, data backup, and index to prevent data leakage.
- Data application layer. The data application layer mainly installs some data security software to prevent the APT attack, user access control, integration tools, and processes. It also realizes system security and ensures data security.
- Data management. The main task is to establish a standard data system, security system, operation system, open system, and authorization management system to ensure security from the perspective of the overall system.

3.3 Data Acquisition Mode

There are four main data acquisition methods, network data capture, data purchase, data exchange, and sharing, commonly applied in the proposed intelligent tourism platform.

3.3.1 Network Data Capture

Internet data capture is to monitor and collect the data of e-commerce websites, social media, and other websites from time to time through web crawler technology. It can crawl any data on the web page, but different collection rules must be adopted for various websites. The data captured by web crawlers include OTA tourists' collection records, attendance records, evaluation, and other forms of data such as heat maps in the search engine, UGC data, and social media data.

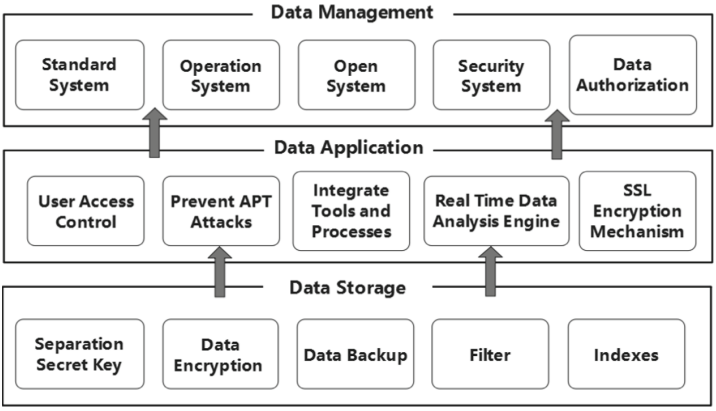

Fig. 3. The security architecture of the smart tourism platform.

3.3.2 Data Purchase

Some traveling data are held in tourism industries and are not disclosed on the internet. These data are necessary for extensive tourism data analysis and management, which can only be obtained through purchase. For example, massive data are generated from mobile communication companies, search engines, and OTA. These data may involve a great deal of user privacy. Therefore, the data purchased and analyzed can only be the data after proper desensitization.

3.3.3 Data Exchange and Sharing

Valuable data held by tourism-related enterprises and departments can also be shared through the exchange. Reliable and efficient data exchange effectively improves tourism service efficiency, reduces operation cost, and enhances competitiveness. Exchanging and sharing data helps break the data island of tourism-related enterprises and departments.

It builds a shared data exchange platform and gives full play to big data's potential value. It also provides timely, accurate, high-quality services for tourism authorities, tourists, and tourism-related enterprises.

3.4 Big Data-Based Smart Tourism Management

3.4.1 Public Sentiment Supervising

The monitoring objects of tourism public opinions include the tourism service departments, distribution centers, scenic spots, tourism principals, etc. Public sentiment supervision and analysis can collect data from websites, WeChat, microblogs, blogs, forums, news websites, post bars, and other data sources. The analysis and processing of the data can realize the functions of public feelings warning, positive and negative emotion analysis, voice trend analysis, data source analysis, public opinion hot words analysis, etc. It also can obtain and analyze the communication process of public opinion hot spots

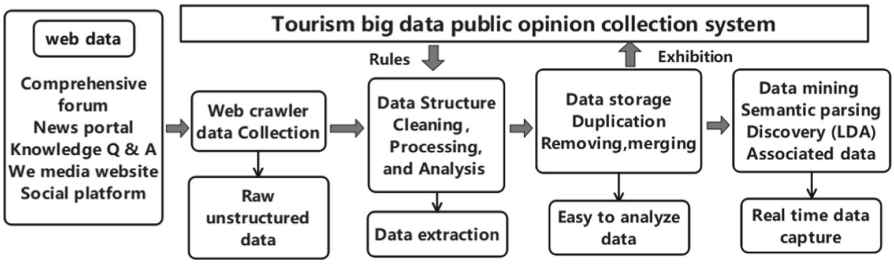

Fig. 4. The process of public sentiment supervising.

in real-time. General feelings' supervision runs through different diffusion, peak, and attenuation stages, deeply observing the event's occurrence, diffusion path, and impact effect, which provides more suitable suggestions in official or unofficial according to the progress. As a result, the competent department quickly builds the ability by using the intelligent platform to grasp social public opinion trends effectively.

What is more, the smart big data-based tourism platform is used to process the tourist comment data of tourists and make a descriptive statistical analysis. The LDA topic model is commonly used to determine the characteristic subject words and their probabilities under the optimal number of topics. Key indicators such as tourism satisfaction and emotional evaluation can guide tourism management with more accurate targets and high efficiency. The primary process of public sentiment supervising is illustrated in Fig. 4.

In addition, a series of prediction techniques, such as the KNN, xgboost, and LSTM, are used to classify the emotional tendencies of tourist comments for the improvement of tourism satisfaction. Finally, with the help of the intelligent tourism platform, the evaluation indicators of tourism satisfaction are classified from the four aspects of tourism landscape, tourism infrastructure, recreation experience, and management service. These indicators are further expanded to establish a complete evaluation index system (natural landscape, regional transportation, tourism experience, and tourism route arrangement) of tourist satisfaction and explore the impact directions on tourist satisfaction.

3.4.2 Tourism Market Status Analysis

The tourism statistics department conducts regular research on the operation of the tourism economy following relevant regulations. The big data service platform can capture and integrate the operation data of hotel accommodation facilities and extensive traffic operation data such as aviation, tourist search, and spatial location data. It provides decision-making reference for the process and investment of urban catering, shopping, scenic areas, and other tourism-related enterprises. In addition, the platform can increase the statistics and data processing capacity in specific operations by innovating the traditional optimization models. It also supports the accurate judgment of tourism enterprises, predicts future trends, and formulates relevant regulatory policies. The platform can also use the composite index method by utilizing tourism big data technology to build and estimate the prosperity index of the tourism market.

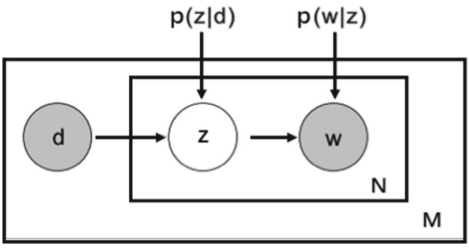

Fig. 5. The calculation of PLSA for tourist portrait management.

3.4.3 Passenger Flow Monitoring in Scenic Spots

The Bluetooth device at the scenic spots grabs the mobile phone signal, obtains mobile operators' data, analyzes the scenic area's statistical data, and other different ways to get the tourist flow data of the scenic spot. Such measures can acquire and store phased historical data. It can also effectively avoid the overload and emergencies of tourist flow in the picturesque area and predict the passenger flow and traffic flow of the scenic spot in the future. Through big data technology, we can analyze the number of inbound tourists, outbound tourists, tourists outside the tourism regions, tourists traveling to other cities in the area, and mutual visits between counties in the counties the town. The RF (random forest), SVR (support vector regression), LSTM (long short-term memory), and PSO-LSTM models are commonly used to predict tourist-related tendencies. With the aid of the above forecasting tools, the platform can reasonably allocate tourism resources, help tourists plan their travel itineraries and improve the tourism experiences.

3.4.4 Tourist Portrait Management

Through the intelligent tourism big data platform, OTA company uses multiple linear regression and probabilistic latent semantic analysis (PLSA) to build tourism consumer portraits based on user preferences, more depth analysis, and mining of user information. The calculation process of PLSA is illustrated in Fig. 5.

The formal representation is shown in Formula 1.

$$p(w|d) = \frac{p(w,d)}{p(d)} = \sum\nolimits_c p(c|d)p(w|c) \tag{1}$$

There w are terms, d is a document, c is topic.

It provides users with more targeted and accurate services, which will help the tourism enterprises develop better. The portrait functions can effectively analyze the users' travel information and precisely predict the tourists' behaviours. Afterward, precise marketing can be carried out for different travel needs and application scenarios. The portrait function also improves the reservation volume of auxiliary ticket products such as visa processing, pick-up service, and foreign currency exchange. In terms of online booking of tourism and holiday products, the smart tourism platform can quickly refine the users' needs through user portraits. The varying requirements of users can suitably specify the family tourism products and their combinations for the consumers of different ages and customized services to meet users' personalized needs. In the field of hotel online

reservation, you can also recommend hotels of different grades according to users' consumption levels and preferences through user portraits to increase the reservation volume of hotels.

4 Conclusions

This research discusses the integration and application of smart tourism and big data. It analyses the construction of the big data-based innovative tourism platform from the aspects of advantages, demand, and benefits combined with the reality of tourism development. The establishment of the intelligent platform is fully discussed from the characteristics of the overall architecture construction, security architecture design, and management functions implementation. In summary, the intelligent tourism platform can better serve the development of the regional tourism industry. Tourists can understand the relevant information on tourism big data through websites, mobile phones, and touch screens. The tourism department can centralize information about tourists, scenic spots, traffic, and other data into the intelligent platform through various tourism service terminals. It facilitates the competent tourism department to quickly and effectively understand the relevant information for correct decision-making, deployment, the realization of the intelligent management for the service of tourism, and paving a new way to promote the sustainable development of creative tourism.

Acknowledgements. This research is supported by the Shandong Province Key R & D Plan (Soft Science) Project (Project ID: 2021RKY07134), and the Institutional Applied Scientific Research Project (ID: 2021yyx-zd02).

References

Chang Xu, 2018. Cross platform We research on the construction of Sanya smart tourism city Based on Internet and Big Data. In ICITBS'18

Colombini, M. (2020). Forecasting tourism demand with multi-source big data. J. Annals of Tourism Research. 502, 1–13.

Haifeng Gao. (2021). Big data development of tourism rresources based on 5G network and Internet of Things system. J. Microprocessors and Microsystems. 80, 1–6.

H Li, T Su. (2002). SM-PageRank algorithm-based user interest model for mobile smart tourism platform. J. Wireless Communications and Mobile Computing. 127, 726–728.

Huang, X., Zhang, L., & Ding, Y. . (2017). The baidu index: uses in predicting tourism flows –a case study of the forbidden city. Tourism Management, 58, 301-306.

Khalid Al Fararni. (2021). Hybrid recommender system for tourism based on big data and AI: A conceptual framework. J. Big Data Mining and Analytics. 4, 47 - 55.

Sen Qiu. (2022). Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges.J. Information Fusion. 80,241–265

Y.Wu, 2021. Application of data mining technology in intelligent pushing urban tourism information system. International Conference on Communication and Electronics Systems (ICCES)