Research on the design of intelligent tourism auxiliary decision-making service platform based on local tourism big data

Zehui Wang[1,a], Dianwei Chi[2*]

[1]School of Physics and Electronic Information, Yantai University, Yantai 264000, China;
[2]Artificial Intelligence Academy, Yantai Institute of Technology, Yantai 264003, China

[a]wangzehuiyt@126.com; *dianwei.chi@163.com

Abstract. With the continuous development of local tourism big data and artificial intelligence technology, it provides the necessary conditions for the construction of local intelligent tourism class auxiliary service and decision-making platform. In this paper, data collection and preprocessing of data from local culture and tourism departments and Internet platforms are carried out, and the design of a smart tourism auxiliary decision-making service platform based on local tourism big data is proposed in combination with artificial intelligence technology. On the one hand, it provides intelligent reference for potential tourists' tourism travel planning and realizes the improvement of personalized services for tourists. On the other hand, it promotes the improvement of the marketing methods of scenic spots, tourist hotels and other business enterprises, and provides auxiliary decision-making support for the tourism management. The design scheme and technical realization scheme of the intelligent tourism decision-making platform formed by this project can be used to guide the construction of the intelligent tourism platform, promote the intelligent transformation and upgrading of local tourism, and promote the high-quality development of local regional tourism.

Keywords: smart tourism; local tourism; big data; assisted decision making; artificial intelligence

1 INTRODUCTION

With the continuous integration of new information technology and tourism, tourism big data is increasingly emphasized. Big data technology is an emerging technology that relies on the gradual development of the Internet, cloud computing and other information technologies [1], and is gradually applied to all aspects of social management. With the rapid development and wide application of cloud computing, big data and artificial intelligence technology, the associated industries of tourism have also accumulated a huge amount of data, which provides data conditions and technical conditions for the construction of a smart tourism assisted decision-making platform based on big data. In the era of smart tourism, more emphasis is placed on the personalization

A. Haldorai et al. (eds.), *Proceedings of the 2024 3rd International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID 2024)*, Atlantis Highlights in Intelligent Systems 11,
https://doi.org/10.2991/978-94-6463-490-7_30

and intelligence of tourism, and it is necessary to explore the potential value in the data, enhance the user experience, provide decision-making support for tourism-related enterprises and departments, and improve the quality of tourism services. Therefore, it is necessary and feasible to build an intelligent tourism decision-making platform that meets local characteristics.

Domestic research on the construction and application of intelligent tourism decision-making service platform based on big data started late, but with the continuous attention to tourism big data and the continuous practical application, the research on big data and artificial intelligence technology in intelligent tourism is increasing. For example, Liang Changyong believes that big data mining is the core of intelligent tourism, tourism can only realize intelligent tourism management, intelligent tourism services, intelligent tourism marketing by deep mining of tourism big data [2]. Wenjun analyzed and mined tourism big data such as spatial location data, consumer usage data, and review data of 32 high-end lodgings, explored the behavioral characteristics and laws of high-end lodging consumers, and provided constructive development strategy suggestions for lodging operators and government departments [3]. Yan Zihan et al. designed an online tourism platform based on big data storage and intelligent recommendation for Guizhou's characteristic towns, realizing the intelligent recommendation of scenic spot information and enhancing the potential attractiveness of the town [4]. At present, Guiyang has constructed a smart tourism service platform based on big data, put forward the theoretical concepts of big data, block data, blockchain and other theoretical concepts, and partially explored and applied the service aspects of big data on smart tourism; Shandong Province Tourism Development Commission completed the construction of a basic collection platform for tourism big data in 2017, integrating the provincial Tourism Development Commission, travel-related units and other data, and initially completing the data grooming involving the key sectors of tourism, access work of data resources.

In summary, there is a gradual increase in the research on the application of big data and artificial intelligence technology in intelligent tourism, but most of them are focused on the application of a certain technology such as intelligent recommendation and sentiment analysis, and there are fewer studies on the overall design and application of the intelligent tourism decision-making service platform. Although some of the studies provide the macro design ideas of intelligent tourism platform, they are not accompanied by detailed design plans and technical realization plans. In addition, there are fewer studies on the current intelligent tourism decision-making platform for building local thematic data warehouse and for local characteristics. Therefore, this study will be based on local tourism data, comprehensively converge the data generated by practitioners and consumers in the local tourism industry, and form a tourism thematic data warehouse that conforms to local tourism characteristics and is oriented to different applications. Then, based on the data warehouse using artificial intelligence technology to achieve emotional tendency analysis, personalized recommendation, auxiliary decision-making rules generation and other intelligent modules, and ultimately form a replicable and implementable local intelligent tourism decision-making platform design and technical implementation of key functions. The design scheme and technical realization scheme of the intelligent tourism decision-making platform formed by this

project can be used to guide the construction of the intelligent tourism platform, promote the intelligent transformation and upgrading of local tourism, and promote the high-quality development of local regional tourism.

2 OVERALL PLATFORM DESIGN

Design an intelligent tourism assisted decision-making service platform based on local tourism big data. The platform functions need to be considered to meet the basic needs of the three core groups of tourists, government and enterprises in the development of the tourism industry, pay attention to the intelligent decision-making assistance of local tourism-related departments, scenic spots and enterprises as well as the personalized experience of tourists, and form a set of replicable and implementable intelligent tourism decision-making platform design scheme. The overall architecture design of the local tourism big data assisted decision-making system is shown in Figure 1:

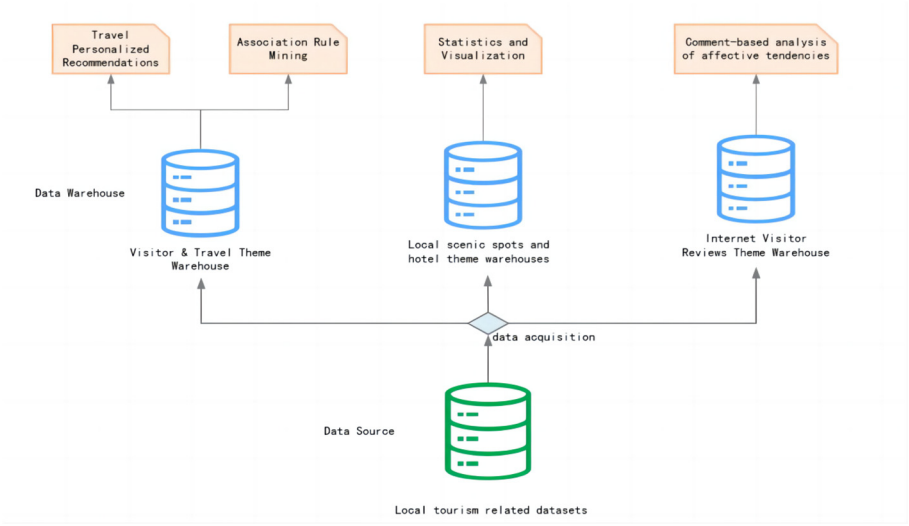

Fig. 1. Overall Platform Architecture

The overall system is divided into three layers: the data collection layer, the data warehouse layer and the data application layer, respectively.

2.1 Data Acquisition Module

Unified collection of tourism-related data in the informatization systems of local tourist attractions, tourism-related enterprises and tourism-related departments, including: data of scenic spot ticketing system, basic data of Internet scenic spots and hotels, and data of Internet tourists' comments. The detailed design scheme and core collection program of the data and information collection function module are formed. This topic collects

local tourism-related data, taking Yantai City as an example, and there are a total of three data sources. First, the basic data of Yantai City Tour Tourist Hotel are crawled from the Internet tourism platform using web crawler technology, including the geographic location of tourist hotels, star rating, tourist satisfaction, type and so on. Secondly, the data of tourists' comments were collected from the Internet tourism platform, including the name of the hotel, tourists' comments, evaluation scores, etc., to provide data support for the emotional analysis of tourists' comments. Third, public tourism data are acquired from the intelligent cultural and tourism big data platform to form a data source that describes the characteristics of scenic area levels and tourist flow decline values, maximum carrying capacity, etc. It provides data support for travel-related departments and enterprises to carry out auxiliary decision support.

The data collected by this module mainly comes from local cities' cultural and tourism big data platforms, Internet tourism platforms or APP platforms. It mainly includes local city scenic spot tourism statistics, Internet platform tourism hotel data, Internet platform tourists' comment data, and tourists' traveling data.

Here, taking local cities in Shandong Province as an example, scenic tourism data can be directly collected from the Yantai Culture and Tourism Big Data platform, and the download format is CSV format, which needs to do data cleaning and feature selection later.

Tourist hotel data and tourist comments and other data are obtained from the Internet platform using web crawler technology, the following is the detailed design of the data collection module.

Here, the design flow of data collection is introduced by taking the collection of tourists' comments and basic hotel data as an example, as shown in Figure 2. User comment data and basic hotel data, tourist travel data, etc. are all from the Internet platform, and this part is realized by using a crawler program. Here, the design process is briefly described as an example of user review data collection of tourism hotels, and the basic data of tourism hotels and tourists' review data are collected from Ctrip.com.

Regarding the data format of hotel user comment data collection, the data data format of the collected data is as follows: the primary key (order_id) of the hotel comment, in order to verify the uniqueness of the information. Hotel webpage address (url), so that it is easy to check whether the data is collected correctly. The name of the hotel (hotel_name), so that it is easy to count the number of hotels. Posting date(post_time) of the hotel review, so that you can know whether it is the most recent posting. The user name of the hotel review (user_name), for user profiling of the review. The content of this hotel review, which is the text data that will be mined later for hotel customer opinions. The user score (user_score) of the hotel review, this data attribute can be mapped to the user review to facilitate the production of the sentiment analysis training set later. The design process of the platform is shown in Figure 1:

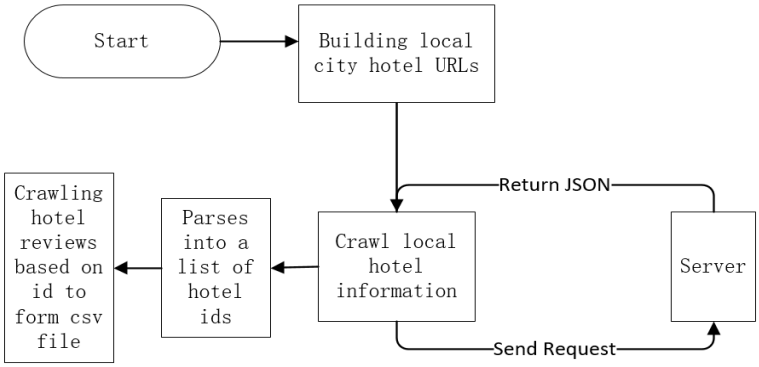

Fig. 2. Data acquisition process

Crawler crawling data need to traverse the collection of tourism hotel links, and then automatically iterated download information, which requires us to analyze and set up the format of the page URL, so that the crawler can automatically traverse the address list to download.

Need to splice the number of pages of tourist hotel reviews, through the tourist hotel ID, review page number can be spliced to complete the review of all tourist hotels JSON address, in the collection of only according to the splicing of the law of iteration can be uninterrupted collection of tourist hotel review data, in the collection of the collection of the time to pay attention to the collection of the speed and the replacement of the IP, to prevent blocking, in order to allow the crawler to uninterrupted operation.

2.2 Data Warehouse Management Module

Based on the collected tourism data, design a data model that meets local characteristics and is oriented to different application needs, form a data warehouse containing different themes, and realize centralized data collection, storage and processing. It provides support for the upper-level big data analysis and decision-making module, and can also provide a unified information platform for information sharing and interconnection among local scenic spots, tourism-related enterprises and tourism-related departments. This topic is based on the three data sources provided by the data collection to design a data warehouse of three topics, i.e., data marts, respectively. They are local tourism hotel basic data theme data warehouse, tourist comments theme data warehouse, scenic spot tourism statistics theme data warehouse.

The volume of collected tourism big data is large, and the data cleaning and preprocessing require large computing resources, so all the data in this module should be considered to be uploaded to the big data platform for storage, which is convenient for data processing with high efficiency, so as to form the final data warehouse table.

Here we take the basic data of tourism hotel and review data theme data warehouse table as an example, and other theme data warehouse table design and realization process is basically the same. The data warehouse table construction process is shown in Figure 3:

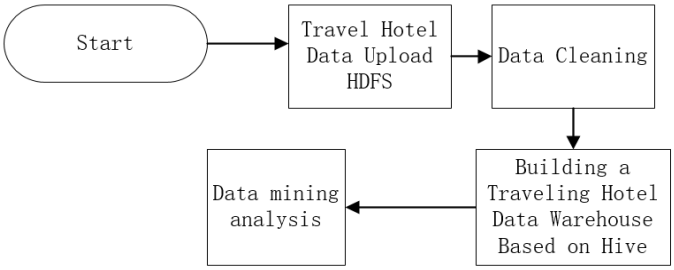

Fig. 3. Data Warehouse Building Process

(1) First of all, based on the basic hotel information data and user review data of a prefecture-level city crawled by the crawler, using Hadoop and other big data platforms to store the data, and uploading the data through Python's hdfs module.

(2) Use Spark and other data analysis frameworks to clean the data set, mainly for abnormal data and inconsistent data.

(3) Based on the storage path of the dataset in hdfs to create hive data warehouse external tables, respectively, the user review data table and the hotel basic information table.

(4) Based on the data warehouse table, according to the dimensions of the hotel and review information that the user cares about, the data analysis and processing is divided into multiple perspectives of concern, and for each of them to create the data warehouse theme internal table, these internal tables can be used to provide a data source for data analysis and visualization in the future.

2.3 Travel Data Analytics and Intelligent Decision Making Module

The key to the local tourism big data-based assisted decision-making platform is the data application level. Data mining analysis based on the data warehouse can enhance the user's tourism experience and provide suggestions and references for improvement and upgrading of scenic spots, hotels and other businesses, as well as provide intuitive data statistics and decision-making support for the relevant governmental management departments, so this module is the key module of this system, which utilizes big data analysis and machine learning algorithmic models. Specific system business functions include: sentiment analysis module, big data analysis and mining auxiliary decision-making module.

1) Sentiment Analysis Module Design

Sentiment analysis technology is able to extract users' subjective emotions and satisfaction level from massive text, which provides an important reference for applications such as opinion mining and online opinion monitoring [5, 6].

Based on the data warehouse table formed by user Internet platform review data, deep learning algorithms are used for sentiment analysis to study whether users are more concerned about the presence of negative emotions, to better utilize online

reviews, and to study customer opinion mining to provide suggestions for the hotel industry.

Since the hotel reviews are more Chinese text, Chinese text semantic flexibility, more difficult to separate words, and the problem of multiple meanings of the word, the project uses a sentiment analysis model based on the combination of BERT [7] dynamic semantic coding and deep neural networks. The model uses BERT pre-training to generate word vectors as model inputs, deep neural networks are used to extract sentiment features, and finally the self-attention mechanism is used to differentiate the importance of the sentiment features in the sentence, so as to improve the accuracy of sentiment classification.

2) Design of Tourism Big Data Analysis and Mining Module

The tourism big data analysis and mining module mainly consists of three parts of the design, and the following is an overview of the specific design.

First, based on the relevant statistical data of local tourist attractions, association rules are generated based on the association rule mining model to provide decision-making support for scenic spots and tourism authorities.

Secondly, based on the basic tourism and hotel data and hotel review data accumulated specifically for the platform, a data warehouse is constructed based on the Hadoop big data platform, and various dimensional data statistics and data visualization presentations are carried out based on the data warehouse to provide intuitive data reference for the hotels and tourism management departments.

Thirdly, based on user travel data and hotel basic data, data mining is carried out to generate association rules to provide decision-making support for travel-related hotel operations and user travel.

The module is mainly based on local tourism data data warehouse of different topics, the use of data mining or machine learning algorithms to generate tourists, scenic spots or travel-related enterprises to provide auxiliary decision-making, the following is based on the statistical theme of tourism scenic spots data warehouse as an example, to introduce auxiliary decision-making mining module detailed design and implementation of the idea of the mining module of the other thematic data warehouse is basically the same idea.

Based on the thematic data warehouse of tourism statistics of scenic spots, mining the correlation relationship between scenic spot level and features such as the decline value of tourist flow, maximum carrying capacity, etc., and searching for high-confidence decision-making reference suggestions.

The data here are from the tourists' visit information of each scenic spot of Yantai City Wisdom Culture and Tourism Big Data Platform, time scope: 2018 to present; space scope: Yantai City; update cycle: October every year; update method: library table, increment.

Some of the table data after preprocessing and feature selection are shown in Table 1:

Table 1. Preprocessed statistics for selected scenic spots

max_load	uid	day_lowers	flow_lowers	scenes_level
49500	10487	0	0	4A
120000	10016	82	71	5A
128841	10025	233	230	5A
37061	10045	123	88	4A
5920	10046	18	18	4A
40000	10048	30	27	3A
10240	10049	9	7	4A
6000	10050	10	10	4A
30000	10051	160	147	4A
21000	10052	0	0	4A
200000	10053	189	163	4A
58183	10054	18	18	4A
101600	10055	37	34	4A

We discretize the columns based on the continuous values of the above data, and such processing can produce rules that are more general and inductive. The data columns involved above are: 'max_load', 'uid', 'day_lowers', 'flow_lowers', 'scenes_level'. We discretize the data in the 'max_load', 'day_lowers', 'flow_lowers' columns into three labeled categories, low, medium, and high, with corresponding category names: low, mid, and high.

The next step is to apply association rule mining algorithm to the data after data cleaning to analyze the correlation between attributes. The technical framework is the Weka framework. The basic process is to save the preprocessed tourist travel sample data into the typical data file format of WEKA platform, load it into the corresponding algorithm processing interface, and configure the relevant parameters for data mining experiments.

The following is based on the preprocessed tourism data to illustrate the specific steps of Weka for big data mining, to find out the correlation rules between tourist satisfaction and other features, the specific steps are as follows:

(1) Prepare data samples. We organize the sample data into arff standard format that Weka software can handle as data samples.

(2) Open the data file using weka software and select the attributes of interest, which will be included in the correlation rules generated in the future.

(3) In the interface that describes the algorithm, select the Associate option to set the parameters such as support and confidence of Apriori algorithm, here set the minimum value of support to 0.5 and the minimum confidence to 0.9, the parameters can be adjusted in this interface.

(4) Click the start button, according to the support and confidence parameters set by the user, the strong association rules are generated, and the generated rules are saved to the database, which is convenient for the merchants and the platform background management personnel to view and export the rules.

The following are the results of the algorithm:

Best rules found.

1. max_load=low flow_lowers=low 7679 ==> day_lowers=low 7678 <conf:(1)> lift:(1.02) lev:(0.01) [131] conv:(66.04)

2. max_load=low day_lowers=low 7679 ==> flow_lowers=low 7678 <conf:(1)> lift:(1.02) lev:(0.01) [118] conv:(59.51)

3. day_lowers=low scenes_level=4A 5265 ==> flow_lowers=low 5264 <conf:(1)> lift:(1.02) lev:(0.01) [80] conv:(40.8)

4.day_lowers=low 9828 ==> flow_lowers=low 9824 <conf:(1)> lift:(1.02) lev:(0.01) [148] conv:(30.47)

5.scenes_level=4A5311 ==> flow_lowers=low 5271 <conf:(0.99)> lift:(1.01) lev:(0) [42] conv:(2.01)

From the above results, we can see many rules that can be used to assist in decision making, for example, the maximum carrying capacity is low and the flow decline is low there is a high probability that the degree of decline is low on a daily basis. Another example is the 4A level scenic area has a 99% probability of a low degree of traffic decline. A scenic spot with a low degree of traffic decline and a level of 4A is certain to have a low degree of daily decline. A low maximum carrying capacity has a low daily flow rate decline, etc.

3 CONCLUSION

In this paper, in the context of the improving tourism big data and the era of intelligent tourism in various places, through the research and design of intelligent tourism assisted decision-making platform based on local tourism data warehouse, and ultimately form an implementable platform design scheme and realization ideas that can guide the development. The overall platform is divided into three layers based on the process of big data analysis, namely, data collection layer, data warehouse layer and intelligent decision-making layer. The detailed design process and key module implementation ideas are given for each layer. The design scheme formed can provide ideas and technical references for local construction of tourism decision-making platform, which is conducive to the integration of local scenic spots, hotels, tourists and other multi-dimensional tourism data, and combined with artificial intelligence technology to carry out personalized recommendations and decision-making suggestions to push, providing decision-making support for tourists' travels and local tourism-related departments and enterprises.

REFERENCES

1. Xiao Yuanping, Gong Xiang. "Internet +" perspective of Guizhou tourism industry wisdom development research[J]. Guizhou Social Science,2016(05):127-132.
2. Liang Changyong,Ma Yinchao,Lu Caihong. Big data mining:The core of smart tourism[J]. Development Research, 2015(5):6.
3. Wen Jun. Research on consumer behavior of high-end lodging based on big data analysis[D]. Zhengzhou University, 2021.

4. Yan Zihan, Yu Zhen, Cao Pan, Wu Chengsi, Wang Jiawei, Zhou Xianjun. Design and realization of online tourism platform for Guizhou characteristic town based on big data storage and intelligent recommendation[J]. Computer Knowledge and Technology,2022,18(24):20-22.
5. HE Yanxiang, SUN Songtao, NIU Feifei, LI Fei. A deep learning model for sentiment semantic enhancement for microblog sentiment analysis[J]. Journal of Computing,2017,40(04):773-790.
6. J. Jiang. Sentiment analysis of social media text[D]. Nanjing University of Science and Technology, 2017.
7. Pang, Liang, Lan, Yanyan, Guo, Jiafeng. Text Matching as Image Recognition[J].2016.