A Study on the Integrated Development of Artificial Intelligence and Tourism from the Perspective of Smart Tourism

Yingying Pei¹, Yingchao Zhang¹,*

¹Sanya University, Hainan Sanya, China, 572000

*Corresponding author e-mail: 193475741@sanyau.edu.cn

Abstract. In order to provide better travel experience for users under smart tourism, a data model that can be used for tourism analysis which is constructed by combining big data, Hadoop and crawler technology. Then, we combine with Item-based recommendation algorithm and the technology collected by crawler technology based on this model. It can provide an example of the specific integration application of big data analysis model and tourism. The results showed that the analysis model can highly recommend scenic spots favored by tourists. So it can provide reference for the integration of ARTIFICIAL intelligence and tourism, as well as a concrete example for the integration of artificial intelligence and tourism in the future.

Keywords: Smart Tourism, Artificial Intelligence, Development of Fusion

The "Year of Smart Tourism" proposed in 2014 has attracted much attention. The so-called "smart tourism" is the application of big data technology and artificial intelligence to the tourism industry, so as to change the traditional tourism service model and promote the high-quality development of the tourism industry. [1] The core of smart tourism is the word "smart", which refers to the combination of new technologies such as artificial intelligence and traditional tourism business. It can recommend interesting tourism information according to the preferences of travel enthusiasts, and reducing tourists' acquisition of tourism information. [2-3]So it can provide a humanized service mode. In the study, we try to combine big data technology to build an artificial intelligence model that can be used for comprehensive tourism analysis. Then, based on this model, taking scenic spot recommendation as a case, construct a tourism and artificial intelligence fusion model that can be used for scenic spot recommendation. It builds an example of integrating tourism and artificial intelligence that can be used for scenic spot recommendation, so as to provide tourists with needed tourism services.[4-5]

1. Basic Technology

1.1. Big Data and Hadoop
Big data technology has been widely used in daily production and life. It has the characteristics of large scale and complex structure, involving structured and unstructured data. It is difficult to effectively manage big data using traditional data management models, so it relies on specialized
storage tools for big data storage and processing. [6-7] At the same time, the big data technology has been applied in many fields with the rapid development, forming different processing algorithms and processing tools. [8-9] Through the processing and analysis of big data, valuable information can be obtained, which helps to provide a scientific basis for business management. Currently, Hadoop is the most typical in the field of big data. The HDFS system can complete the storage of massive data, which makes up for the deficiencies of traditional management tools. The high-performance MapReduce parallel computing framework is popular in data processing. The framework is divided into two parts: Map and Reduce. Map is used for data reading and processing and Reduce is used to complete the processing according to the set steps and summarize all the results. Based on the cooperation of the two to complete the process of massive data processing, it can provide certain decision support.[10]

1.2. Artificial Intelligence and Python
Artificial Intelligence (AI) technology has become a popular research field. The application of this technology has enabled machines to gradually possess the characteristics of human thought and behavior, enabling the machine to make judgments and processing on its own according to the acquired information and perform. This reduces the pressure of manual operation. In AI, there are mainly two modules: deep learning and machine learning. In the existing research, many algorithms have been formed and applied to the actual production field. The Python language is widely used in artificial intelligence, and mature development libraries such as Sklearn and Numpy have been formed, which can be directly used to process data and analysis.

Part of data in this research was obtained directly from the enterprise. In addition, crawler technology was used to obtain data from the network. The main libraries we used are BeautifulSoup, Requests, etc. After obtaining enough data, the process of deep learning can be carried out, relying on specific algorithms for training.

2. Intelligent Tourism Big Data Analysis Model Based on Artificial Intelligence

2.1. Overall Framework
In the process of the continuous improvement of my country's urbanization, people's income levels have increased, and more new needs have emerged on the spiritual level. Many residents choose to travel to relax themselves, cultivate their sentiments, and seek to return to nature. In this situation, the tourism industry has obtained a greater opportunity for development. With the continuous development of the tourism industry, the needs of travel enthusiasts become diversified so that it is necessary to provide traveler targeted services according to different travel needs. Therefore, it is necessary to combine big data and other technologies to analyze the preferences of tourists. So we can innovate in tourism products and services. In addition, we can optimize the allocation of tourism resources and provide consumers with higher-quality tourism services. Based on the above crawler technology and Hadoop technology, the smart tourism big data model established in this research is shown in Figure 1.
2.2. Model analysis
According to Figure 1, the smart tourism data model constructed in the study is divided into multiple levels. This approach based on layered architecture enhances the logic and scalability of the system.

(1) The hardware resource layer belongs to the basic layer, creating basic hardware conditions for the normal operation of the entire system.

(2) The data processing layer is mainly related to the data preprocessing. On the one hand, it is necessary to prepare basic data, that is, to obtain data from the network through Python crawler technology, or to collect existing data in the enterprise. On the other hand, it preprocesses the acquired data including data cleaning and other processes, and lays a data foundation for the subsequent learning and processing process.

(3) The data analysis layer belongs to the key process of data processing. In this process, it is mainly processed through the Numpy library and the Sklearn library.

(4) The artificial intelligence layer belongs to the core level, which combines specific tourist areas to establish a learning model, uses the training set to complete the testing process, and optimizes parameters.

(5) Application service layer
Combining the deep learning model obtained by training for predictive analysis, providing personalized tourism and user demand analysis services, which can meet the needs of different travel enthusiasts and improve the quality of tourism services.

(6) The entire model
The overall model is divided into two levels, namely data modeling and basic cloud services. The latter can effectively improve the efficiency of the calculation process.

3. Analysis Model Examples
Based on the above model construction, this article uses scenic spot recommendation as the entry point of artificial intelligence analysis. And then we build an analysis model for scenic spot recommendation. Specifically from the following aspects.

3.1. Introduction to Item-based Attractions Recommendation
At present, different attractions have formed in many areas. Tourist enthusiasts have different
preferences in attractions. Some tourists are keen on historical and cultural attractions, and some prefer various types of natural landscapes. So they have significant differences in the choice of scenic spots. Many tourists express their attitudes and opinions through the scores of scenic spots, and maintain a high consistency with their personal preferences. Different tourists often score differently for different scenic spots. Based on the model in Figure 2, the model in the artificial intelligence layer is designed as an Item-based recommendation algorithm. According to the user’s rating, the tourist’s preference information is obtained and then processed by the algorithm which recommends similar attractions to tourists, enabling them to obtain information on the attractions of interest more quickly. It provide a personalized service model for travel enthusiasts to meet different needs in travel. The specific recommended principle is shown in Figure 2.

![Figure 2. Schematic diagram of Item-based attraction recommendation algorithm](image)

During the execution of the algorithm, a score matrix needs to be constructed first. In detail, a "tourist-spot" matrix is formed according to the scores of all scenic spots. According to the information in Figure 1, there are certain rules in the preferences of tourists for different scenic spots. Among them, tourists who prefer scenic spots I1 tend to show higher interest in scenic spots I5, so the two are more similar. Therefore, when the tourist U4 selects one of the above two scenic spots, we can recommend another scenic spot to him. So we can quickly recommend the scenic spot of interest to the user.

3.2. Algorithm Flow

The specific steps recommended by Item-based are:

Input: The input data mainly include tourist U, "visitor-attraction" rating set R, and nearest neighbor K.

Output: Output the attractions that users may be interested in which is the recommended results.

Step 1: First construct the "tourist-scenic spot" rating matrix.

Step 2: Secondly calculate the similarity. In detail, calculate the similarity between each scenic spot and the user's unrated scenic spot, and select the top K according to the order of similarity from largest to smallest. The specific calculation formula is as follows.

\[
sim(i, j) = \frac{\sum_{w \in \mathcal{U}} (R_{u,i} - \overline{R}_i)(R_{u,j} - \overline{R}_j)}{\sqrt{\sum_{w \in \mathcal{U}} (R_{u,i} - \overline{R}_i)^2} \sqrt{\sum_{w \in \mathcal{U}} (R_{u,j} - \overline{R}_j)^2}}.
\]

In the formula, \( R_{u,i} \) and \( R_{u,j} \) respectively represent the scores of tourist U for scenic spots i and j, and \( \overline{R}_i \) represents the average score. According to the above formula, the Pearson correlation coefficient can be calculated but a problem needs to be considered in the calculation process which is how to avoid the denominator being equal to zero. So targeted design is required. If the scores of two scenic spots are the same during the calculation process, and the denominator appears to be zero, then the value \( \sim(i, j) = 1 \) is taken. And if the denominator is equal to zero because there is only one tourist rating, the value \( \sim(i, j) = 0 \) is taken.
Step 3: Combining the above process to predict, the predicted score value of the unrated scenic spot can be obtained. The specific formula is as follows.

\[
R_{u,i} = \frac{\sum_{j \in N} \text{sim}(i, j)R_{u,j}}{\sum_{j \in N} \text{sim}(i, j)}
\]

In the above formula, \(R_{u,i}\) and \(R_{u,j}\) represent the tourist's predicted score and historical score for scenic spots i and j, respectively. And \(\text{sim}(i, j)\) represent the similarity of the two scenic spots i and j.

Step 4: Provide the final recommendation results to tourists.

3.3. Analytical Model Verification

3.3.1. Data Sources
In this study, data from Ctrip.com was used. This website mainly provides consumers with travel services, including scenic spot inquiry, ticket booking, and scenic spot rating. Based on the information in this website, the required basic scoring data can be obtained. Tourists need to rate the scenic spots. The indicators of the score include cost-effectiveness, interest, and scenery, combined with their own feelings (between 1-5 points). Based on the crawler technology, the score data of Hainan's scenic spots is obtained, and then the data is organized in a certain format. The first two columns are the scenic spots and the user names, and the following three columns correspond to the scores of the three indicators. After screening and other processing, the final score data has 133864 conditions, in which the numbers of users and scenic spots are 32587 and 343 respectively.

3.3.2. Result Analysis
Among the three scoring indicators, the “scene” indicator is selected first, and the data is divided into two parts: the training set and the test set. The numbers of the two are 100000 and 33864, respectively, and \(k=10\). The final calculation result shows that \(\text{MAE}=0.696\), which verifies the effectiveness of the scenic spot recommendation algorithm designed in this paper.

4. Conclusion
The traditional tourism service mode has been difficult to meet the personalized tourism demand. Traditional methods can not integrate tourism resources. It only rely on a small amount of data to design tourism products and services and can not meet the needs of tourists in terms of tourism routes, tourism methods, and can not recommend the tourism information that is really interested in. The recommendation system based on artificial intelligence has certain advantages. It can process and analyze tourists' preference information, and provide users with satisfactory tourism information and services. So it can provide necessary support for the high-quality development of tourism industry. In this study, the smart tourism big data analysis model is constructed and the artificial intelligence technology is introduced, which can recommend the scenic spot information of interest for tourism lovers. So users can obtain high-quality tourism experience in real time.

References

[1] Yang Liu, Lin Xiaohua (2020). Development of tourism location recommendation system. Information technology and informatization, no. 11, pp: 13-15.
[2] Wang Yongcai (2020). Data feature mining method of electricity consumption behavior based on Hadoop platform. Automation and instrumentation, no. 11, pp: 227-230.
[3] Yang Ping (2020). Yunnan ethnic culture tourism information recommendation system based on cloud server. Microcomputer application, vol. 36, no. 11, pp: 33-35
[4] Pan Yun (2020). Information construction and improvement of sanatorium based on Hadoop technology. Life science instruments, vol. 18, no. 5, pp: 59-65
[5] Gao Jialiang, Qiu Peiyuan, Yu Li, et al. (2020) Interpretable scenic spot recommendation based on tourism knowledge map [J]. Chinese Science: Information Science, vol. 50, no. 7, pp: 1055-1068

[6] Zheng Songyin, Tan Guoxin, Shi Zhongchao (2020). Research on tourist attractions recommendation model based on segmented user groups and time context. Data analysis and knowledge discovery, vol. 4, no. 5, pp: 92-104

[7] Zhou Xiong, Zhou Zemin, Peng Yanjun, et al. (2020) Time frequency signal discharge recognition algorithm research and field application of characteristic atlas. Electrical technology, vol. 21, no. 6, pp: 63-68

[8] Liao Wangyu (2020). Personalized travel route recommendation based on travel notes graphic information [J]. Computer applications and software, vol. 37, no. 6, pp: 108-113

[9] CHEN Si, Tian Jingyang (2020). Research on the model of scenic spot recommendation based on collaborative filtering algorithm. Modern electronic technology, vol. 43, no. 11, pp: 132-135

[10] Jeong Chi Seo, Ryu Ki Hwan, Lee Jong Yong, Jung Kye Dong (2020). Deep Learning-based Tourism Recommendation System using Social Network Analysis. International Journal of Internet, Broadcasting and Communication, vol. 12, no. 2.