Application of big data technology in smart Tourism

Nan Chen1,*
1NanChang Institute of Science and Technology, Nanchang, 330108

*Corresponding author e-mail: 49970707@qq.com

Abstract. With the improvement of information technology, big data thinking has penetrated into all aspects of our social life and business fields. With the development of economy, tourism has become the pillar industry of the national economy. In order to speed up the development of tourism industry, many provinces and regions in China also follow the pace of big data era and make full use of big data technology to develop smart tourism. In today's information age, big data technology has important application value and significance for the development of smart tourism. Based on this, this study aims to study the application of big data technology in smart tourism. This paper mainly uses data mining technology and Bayesian network model analysis technology to analyze the big data of travel ticket data of smart tourism as an example to study the application of big data technology in smart tourism. It is found that the Bayesian network model can be used to predict the future trend of air ticket price changes by using the data of a certain period of time in the past, and the prediction accuracy rate is more than 80%, which is more accurate than the traditional neural network model data analysis technology. In addition, big data technologies such as data mining have been widely used in smart tourism, which greatly improves the intelligent degree of tourism industry, and has important significance for the development of smart tourism.

Keywords: Big Data, Smart Tourism, Bayesian Network Model, System

1. Introduction

1.1. Background and Significance

With the development of Internet, Internet of things, Internet of vehicles and big data technology, today's society has entered the era of big data. Whether in social life or business field, data is everywhere. In today's era of big data, data is regarded as a huge "treasure house". The application of big data technology to fully mine and analyze the massive data generated by Internet, Internet of things, Internet of vehicles and other networks can bring huge social benefits and considerable economic benefits to the country and society. Thus, data has become an important factor of production in the era of big data. With the continuous growth of data today, the huge volume, diverse types and different values of data bring us both opportunities and challenges. As a tool to explore and use this huge treasure house efficiently and reasonably, the application of big data technology in the field of tourism also attracts people's attention and discussion.
With the improvement of people's living standard and the change of life concept, tourism seems to have become an indispensable part of people's life. At present, all provinces and cities in China attach great importance to the development of tourism. Coastal cities such as Hainan are also actively building international tourism islands to promote the development of tourism. The rapid development of tourism industry has accumulated a large number of tourism data. These huge and diverse data not only bring us huge economic benefits, but also bring us new challenges in how to deal with these data. The application of big data technology to the analysis of tourism data can make more effective use of tourism data resources, create greater economic and social benefits, and further promote the development of tourism. Throughout the world, in the era of big data, tourism has begun to realize the importance of big data. Data collection and analysis through big data technology can provide decision-making reference for the development of tourism. For example, during the 2012 London Olympic and Paralympic Games, ecircle, a US company mainly engaged in data analysis, used big data technology to help Londoners plan their travel arrangements through specific websites to avoid people flocking to hot spots at the same time. In China, many provinces and autonomous regions also comply with the trend of big data era and vigorously develop smart tourism.

At present, many areas in China have carried out the construction of smart tourism. In 2010, Jiangsu province put forward the concept of "smart tourism" and explored in Nanjing, Zhenjiang and other cities. This is an important attempt to develop tourism with data processing technology, and has achieved good results. Different from the traditional information system construction, "smart tourism" pays more attention to intelligent technology and humanized service. Its purpose is to enable users to meet the needs of tourism information application services anytime, anywhere. This form of "smart tourism" involves a variety of modern science and technology, extending and expanding the existing tourism information construction, striving to build a huge intelligent information service system. For example, Travelocity, an online travel website, uses big data analysis in pricing, inventory and advertising, data mining, best transaction analysis and recommendation, in order to provide the right tourism products to target customers and improve customer satisfaction. However, the traditional data processing technology is often unable to adapt to the increasingly large volume and type of tourism data, so the application of big data technology in smart tourism is more valuable and meaningful.

1.2. Innovation of This Paper

In this study, how to integrate big data thinking into the user experience design of smart tourism products is taken as the research object. It is of great significance in theory and practice to creatively combine big data thinking, smart tourism and user experience into the construction and development of smart tourism. From the theoretical level, first of all, the mass of big data makes the user experience elements in smart tourism products more complete. Big data replaces the traditional sample instead of all statistical methods, and uses global data to analyze the results more accurately and close to the truth. Abandoning the exploration methods with subjective judgment in the past and presenting the facts with data can help us analyze the behavior characteristics and motivation of tourists, so as to refine the key needs of user experience. Secondly, based on the evaluation dimensions and indicators of user experience of tourism products, a referential evaluation model of user experience quality of smart tourism products is established. Finally, based on a solid theoretical foundation and a large number of case studies, this paper summarizes a set of application methods and approaches of big data thinking in the user experience design of smart tourism products.

From the practical level, after a series of calculation and statistics, big data technology is easier to find the development law of things. The relevant theories and methods of this study are helpful for tourism enterprises to formulate corresponding sales strategies, carry out precision marketing, select the most appropriate solution, attract more users and help scientific decision-making. And change people's thinking mode based on experience in the past, explore the great value of big data, and after the new data analysis method, master certain rules and new thinking mode, and predict the future, which is conducive to risk control. Finally, the method summarized in this study is helpful to improve the user experience of smart tourism products. For tourists, it can greatly improve the tourism
efficiency, save time and energy, maximize the satisfaction of demand, and really benefit from tourism.

2. Application of big data technology in smart Tourism

2.1. Big Data Technology

With the development of digital information technology and the arrival of cloud era, more and more people begin to pay attention to big data. Most analysts say that the amount of unstructured data and semi-structured data generated by the creation of a company is huge and of various types, which can not be described by traditional data, and big data is also born. If these data are downloaded to the relational database for analysis, it may cost too much resources. Therefore, people pay more and more attention to the application of big data technology. Generally speaking, real-time big data set analysis needs a framework similar to MapReduce to allocate work to a large number of computers. Therefore, people usually associate cloud computing with big data analysis. To process large amounts of data effectively requires special big data technology, which can be tolerated over time. The strategic significance of big data technology lies in specialized processing of meaningful data from massive data. In other words, if big data is an industry, the key to the industry's profitability is to improve the "processing capacity" of data and realize the "value-added" of data through "processing".

2.2. Smart Tourism

The concept of "smart tourism" comes from "smart earth". The core idea of "smart earth" is to change various interaction modes in society in a more intelligent way by using the new generation of information technology, so as to achieve the accuracy and efficiency of information interaction. "Smart tourism" is an important part of the smart earth. With the help of modern information means, it constructs a complete intelligent tourism service system for the development of regional tourism industry. The system is oriented to the government, enterprises and tourists. It provides more intelligent and diversified information services on the basis of digital city, and uses more convenient terminals to build a complete wisdom for tourism activities Can solve the problem. At present, the elaboration of "smart tourism" focuses on the level of smart technology and smart terminal. Of course, intelligent modern technology is the proper meaning of "smart tourism", but from the perspective of system engineering, technology and terminal are not the only components of "smart tourism". As we all know, data is the source of information services. The primary goal of building "smart tourism" should be the construction of smart data.

2.3. Intelligent Tourism Information System (ITIS)

The concept of Intelligent Tourism Information System (ITIS) is derived from tourism information system (TIS). It is an important part of Intelligent Tourism System and the concentrated embodiment of "service side" in the construction of intelligent tourism system. Based on the data center, itis mainly provides various information services for users through platform construction. The information services provided by its mainly include: tourism resources promotion, tourism shopping, tourism itinerary design, tourism traffic inquiry, tourism department management, scenic spot management, etc. As the "server" construction part of "smart tourism" system, itis also the key part of "smart tourism" to play a role, and also the key to information services. Under the system of "smart tourism" integrated with a variety of technical means, its also makes full use of modern information technology, applies GIS, virtual reality, video monitoring and other technologies to the construction of the platform, and strives to build an integrated information service system that meets the needs of users and can provide various information services under the overall framework of "smart tourism".

2.4. Bayesian Network Model

Bayesian network is composed of a directed acyclic graph and conditional probability table. It describes the dependence between variables by directed acyclic graph. It is a commonly used
probability graph model for big data analysis. In Bayesian network model, the directed edge between nodes is usually used to represent the conditional dependence and independent relationship between variables, such as [23,24], and each node represents a variable. If the arrow in the graph points from node A1 to A2 and A3, A1 is said to be the parent of A2 and A3, otherwise A2 and A3 are said to be the children of A1. In addition, in Bayesian networks, each node has a conditional probability distribution associated with it, which is represented by cpd function. Given the state of the parent node, the conditional probability table of variables with different values is represented by cpd function, and the strength of the relationship between variables is expressed by conditional probability. If the variable has no parent node, the information is represented by a priori probability.

The common methods of Bayesian network structure learning are based on score search. The scoring search method first defines a scoring function as a measurement standard to score all the structures in the model structure space, and uses the search algorithm to select the network with the highest score as the final network structure. The main scoring functions based on Bayesian statistics are K2 score, BDE score and BIC score. In BN structure learning, structure n and parameter α are regarded as random variables in Bayesian network composed of m variables a = {A1, A2, ···, am}. Let AI have ti values and its parent node PA (AI) have PI values. Parameter α = {α ijk| I = 1, ···, M; J = 1, ···, PI; k = 1, ···, Ti}. For a given data set s, assume the network structure. The structure of the network can be obtained from the Bayesian formula, and the maximum posterior probability of the model can be obtained:

\[
P(N / S) = \frac{P(S / N) * P(N)}{P(S)}
\]

(1)

P (s) has nothing to do with the network structure n. Generally, the maximum value after logarithm is taken as the maximum value of the formula, that is, \(\log P(n, s) = \log P(s | n) + \log P(s)\), which is defined as the Bayesian score of network structure. Suppose that the data set s obeys uniform distribution and contains complete independent and identically distributed data of samples \{A1, A2, ···, an\}

\[
F_{k2}(N / S) = \sum_{i=1}^{m} \sum_{j=1}^{p_i} \log \frac{(t_i - 1)!}{(n_{ij} + t_i - 1)!} + \sum_{k=1}^{t_i} \log n_{ikj}
\]

(2)

Where nijk represents the number of samples when the variable AI takes the kth value and PA (AI) takes the jth value. If s and N satisfy the above hypothesis and P (α| n) obeys the Dirichlet distribution, the corresponding BD (Bayesian Dirichlet) score is obtained:

\[
F_{BD}(N / S) = \sum_{i=1}^{m} \sum_{j=1}^{p_i} \left[ \log \frac{\Gamma (\beta_{ij})}{\Gamma (\beta_{ij} + n_{ij})} + \sum_{k=1}^{t_i} \log \frac{\Gamma (\beta_{ijk})}{\Gamma (\beta_{ijk} + n_{ijk})} \right]
\]

(3)

Where \(\beta_{ijk}\) is a parameter in the Dirichlet distribution. When \(\beta_{ijk} = 1\), BD score and K2 score are the same. In addition, if the prior information distribution of structure obeys uniform distribution, the corresponding scoring function is consistent with that of BDE

\[
F_{BDE}(S / N) = \sum_{i=1}^{m} \sum_{j=1}^{p_i} \left[ \log \frac{\Gamma (\frac{a}{p_i})}{\Gamma (\frac{a}{p_i} + n_{ij})} + \sum_{k=1}^{t_i} \log \frac{\Gamma (\frac{a}{p_i} + n_{ijk})}{\Gamma (\frac{a}{p_i})} \right]
\]

(4)

3. Experimental Experiments
3.1. Subjects
The research object of this experiment mainly obtains the ticket price data stored in SQL database, such as flight number, class status, collection date, departure date, departure city and arrival city, and through Bayesian network model analysis of these data to predict the trend of travel ticket prices in the next few days.

In this experiment, three scoring functions based on Bayesian network model are mainly used to analyze the travel ticket price data, and predict the change trend of the ticket price in a few days, so as to obtain the prediction accuracy. It can be seen that the Bayesian network model is more accurate than the traditional data analysis technology, which is helpful for us to learn the Bayesian network model. This experiment is divided into experimental group, control group and analysis group.

In the experimental group, the researchers collect the experimental data obtained by the experimental personnel through the intelligent tourism information system, and then use the data processing technology to collect and preprocess the data for the analysis of the later experimental conclusions. The comparison group and the analysis group respectively processed and scored the data through different data analysis techniques. The advantages of Bayesian network model analysis technology are found out from the differences between travel ticket price data and the accuracy of prediction results. Finally, the analysis group makes a whole and detailed analysis of the information and data obtained in the whole experimental process, and draws the most accurate and rigorous conclusion.

3.2. Experimental Design
In the process of this experimental study, on the basis of referring to a number of relevant literature and some existing smart tourism information, combined with its own research purposes, this paper designs an experimental scheme for the application of big data technology in smart tourism. There are four themes in the experimental scheme. One is to obtain and preprocess the air ticket price data through the intelligent tourism information system, and collect and classify the data to get useful data. Secondly, the Bayesian network model is established to integrate and analyze the obtained data information to find out the useful information for the experimental research, and further analyze and predict the change trend of air ticket price. The third is to use the scoring method based on Bayesian network model to calculate and analyze the data obtained by using K2, BDE and BIC scoring functions respectively, and compare the analysis and prediction accuracy of several functions. Finally, the paper makes an overall analysis of all the useful data information obtained, and concludes the shortcomings in the application research process of big data technology in smart tourism. This paper discusses and studies these deficiencies and tries to overcome them and eliminate them.

3.3. Statistical Methods
SPSS13.0 was used to analyze the experimental data, LSD method (homogeneity of variance) and dunnetts method (non-uniformity of variance) were used for variance analysis, and the test level was $\alpha = 0.05$.

4. Discussion

4.1. Tourism Ticket Data Mining
The main data collected in this experiment are the comparative data records of ticket prices before and after each flight of different routes. The collected data of ticket prices of scenic tourism are deeply mined. The data mining method is used to find out the characteristics of ticket price changes, so as to provide more scientific travel guidance for tourists. For the data of travel ticket price, the data mining method is used to calculate and analyze the data scientifically, and the accurate ticket price data is finally obtained. The data can provide reference for tourism activities and provide more humanized information service for tourists.
### Table 1. Advance booking time

| Air Route       | Airlines          | Advance booking time |
|-----------------|-------------------|----------------------|
|                 |                   | One day  | Three days | Seven days | Fifteen days | Thirty days | Sixty days |
| Beijing | Guangzhou         | AIR China | CT RIP | ¥15  | ¥17  | ¥14  | ¥15  | ¥15  | ¥13  | ¥15  | ¥12  | ¥17  | ¥10  |
|                |                   | HN       | CT RIP | ¥12  | ¥14  | ¥12  | ¥12  | ¥10  | ¥12  | ¥13  | ¥14  | ¥16  | ¥17  |
| Beijing | Shanghai          | AIR China | CT RIP | ¥17  | ¥18  | ¥17  | ¥17  | ¥17  | ¥17  | ¥17  | ¥17  | ¥17  | ¥17  |
|                |                   | HA       | CT RIP | ¥16  | ¥16  | ¥15  | ¥15  | ¥14  | ¥14  | ¥15  | ¥14  | ¥17  | ¥17  |
| Beijing | Shanghai          | HA       | CT RIP | ¥16  | ¥16  | ¥16  | ¥15  | ¥14  | ¥14  | ¥16  | ¥14  | ¥16  | ¥16  |

#### 4.2. Analysis of Bayesian network prediction results

The prediction results will be analyzed from the following three aspects. First of all, the existence of residual seat information is discussed, that is, whether residual seat information should be included in the data is compared; then, we use search scoring algorithm to select BIC respectively. By comparing and analyzing which scoring function has higher or more stable prediction accuracy, the network structure and prediction accuracy of different nodes are compared, and the more appropriate number of nodes is selected to guide the research of practical problems. All machine learning training sets select the sample as the first 70% of the total evidence, and select the remaining 30% as the corresponding test set. Using predict function, we can use the probability of network structure and condition calculated by machine learning, take the variables of the first k - 1 network node in each test data set as the variable in the total evidence of training set, and predict the probability that the k th node will change into upward, invariant and descending. The class with the highest probability is taken as the prediction result of the K node, and the prediction accuracy of a, B, C is calculated by using the table function to compare the predicted value with the real value, and the prediction accuracy of a, B, C is calculated by comparing the predicted value with the true value by the table function. The relative accuracy is calculated as follows: The total accuracy rate = the number of correct prediction / the number of prediction samples; The accuracy rate of a = the number of correct a / the number of a in the prediction set; The accuracy rate of B = the number of correct predicted B / the number of B in the prediction set; The accuracy rate of C = the number of correct predicted C / the number of C in the prediction set.
prediction set; The accuracy rate of C = the number of correct predicted C / the number of C in the prediction set; Finally, the accuracy of each node's three scoring functions is saved in the accy dataset to reduce the number of nodes.

Figure 1. Total prediction accuracy of three scoring function

Figure 1 and Figure 2 show the trend of the overall prediction accuracy of the three scoring functions and the prediction accuracy of B. as shown in Figure 1, the overall accuracy decreases slowly with the increase of the number of nodes. The change trend of BDE grading function and BIC grading function is the same, but K2 grading function fluctuates greatly.

Figure 2. The prediction accuracy of B for three scoring function

It can be seen from Figure 2 that the prediction accuracy of the three scoring functions for B varies with the change of nodes. BDE is relatively stable, BIC is relatively unstable, and K2 scoring function is the most unstable. According to the score of network structure diagram, we can see that the network structure diagram trained by K2 score function is also very complex, and the function and network diagram method using BIC and BDE are relatively simple, and BDE network is the simplest. According to the score of network structure chart, it is found that there is no significant difference in the score of network structure among the three scoring methods. Considering the above considerations, we decided to select the BDE network structure scoring method to train the network structure. Therefore, the following research is mainly based on BIC and BDE scoring function and network diagram.

5. Conclusion

With the rapid development of tourism, the increase of the number of tourists and the fierce competition in the tourism industry, these have brought great difficulties to the development of local tourism industry, the planning and layout of tourism facilities, the development direction of transportation and other tourism decision-making. On the other hand, a large number of tourism data accumulated for a long time can not be used. Therefore, this paper applies the Bayesian network
model analysis technology of big data technology to the tourism industry, sorts out and analyzes the massive data accumulated in the tourism industry, finds out the hidden knowledge that has a significant impact on the tourism industry, so as to provide scientific basis for the management and decision-making of the Tourism Management Department, and promote the development of the tourism industry. The application of big data technology in smart tourism is not only more intelligent in tourism management and more effective in the use of tourism resources, but also can increase the satisfaction of tourists in tourism service, which has great research value and application prospects.

References
[1] Liu Z , Wang Y , Cai L , et al. Design and manufacturing model of customized hydrostatic bearing system based on cloud and big data technology[J]. International Journal of Advanced Manufacturing Technology, 2016, 84(1-4):261-273.
[2] Zhang Z , Barbary K , Nothaft F A , et al. Kira: Processing Astronomy Imagery Using Big Data Technology[J]. IEEE Transactions on Big Data, 2016:1-1.
[3] Ma J , Cheng J C P . Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology[J]. Applied Energy, 2016, 183:182-192.
[4] Lambrou M . Innovation Capability, Knowledge Management and Big Data Technology: A Maritime Business Case[J]. International Journal of Advanced Corporate Learning, 2016, 9(2).
[5] Sang H , Kyoung K , Eun C , et al. System Framework for Cardiovascular Disease Prediction Based on Big Data Technology[J]. Symmetry, 2017, 9(12):293.
[6] Koman G , Kundrikova J . Application of Big Data Technology in Knowledge Transfer Process between Business and Academia[J]. Procedia Economics and Finance, 2016, 39:605-611.
[7] Peng L I , Xiaojun Z , Liang M , et al. Application of big data technology in operation analysis of catalytic cracking[J]. Chemical Industry and Engineering Progress, 2016, 35(3):665-670.
[8] Sledgianowski D , Gomaa M , Tan C . Toward integration of Big Data, technology and information systems competencies into the accounting curriculum[J]. Journal of Accounting Education, 2017, 38:81-93.
[9] Luis Emilio Alvarezdionisi. Envisioning Skills for Adopting, Managing, and Implementing Big Data Technology in the 21st Century[J]. International Journal of Information Technology & Computer ence, 2017, 9(1):18-25.
[10] Wei-Chih C , Wen-Hui C , Sheng-Yuan Y . A Big Data and Time Series Analysis Technology-Based Multi-Agent System for Smart Tourism[J]. Applied Sciences, 2018, 8(6):947.