Construction of Tourism E-commerce Marketing System Based on Data Mining

Li Tan[1]

[1]170610749@qq.com

[1]Chongqing College of Architecture and Technology, Chongqing, 401331

Abstract: In order to develop the great potential of tourism e-commerce, increase the market share of tourism e-commerce, thus promoting the development of China's tourism industry and driving the GDP growth of China's tourism industry. By using Java language, J2EE framework and association rule algorithm technology, this paper proposes to mine knowledge from massive tourism information, and extract useful data from it, so as to help China's tourism e-commerce platform analyse marketing strategies, help it better serve the public, meet the current needs of consumers, and provide a basis for enterprises to open up markets and make targeted marketing decisions in the future, thus improving the service quality and enterprise benefits of tourism enterprises.

Keywords: data mining; Java; e-commerce marketing; tourism; electronic information

1 Introduction

With the deepening of economic globalization, the accelerating of social development, the continuous improvement of living standards, and the improvement of spiritual needs after the material needs are met, Chinese residents begin to seek different ways to relax, and the consumption structure of residents is constantly being upgraded and optimized. More and more people regard tourism as the main way of entertainment and leisure. Bathing in the spring breeze of all-round development in the new era, the tourism industry is striding over the "new normal" and welcoming the changes and challenges of the global integrated economy. Under the premise of structural adjustment in many traditional industries, the tourism industry is also facing new opportunities for innovation and development, requiring the development speed to be accelerated and the development scale to achieve new breakthroughs. With the advent of the information age, the number of Internet users in China has increased by leaps and bounds, and the integration of tourism and e-commerce has become an inevitable trend. The development of e-commerce and network marketing has given tourism e-commerce a platform and opportunity for development. According to the analysis report on the development of China's online tourism market released by China Commercial Industry Research Institute, as of December 2019, the number of tourists in China has reached an astonishing 2.01 billion person-times, of which 560 million people use online network platform to pay for tourism projects [5]. It can be seen that under the double impetus of the continuous expansion of the tourism market and the wide

ICIDC 2022, June 17-19, Qingdao, People's Republic of China

DOI 10.4108/eai.17-6-2022.2322636

application of information technology, the subject of tourism e-commerce has developed rapidly, but overall, the penetration rate of online tourism in China is still at a low level, and tourism e-commerce has great development potential and market prospects. With the passage of time, the tourism industry has a wide range of points, and the characteristics of mobility and scale are very obvious. The data generated by these tourism e-commerce platforms show an increase of geometric magnitude, which can be called information explosion [9]. Faced with such a huge amount of data and information, we need to consider how to better handle the data and information, find useful data in these data and information, and quickly discover and utilize them, so as to further improve the use efficiency of data and information, and make them truly valuable resources for users. At present, the construction mode of tourism e-commerce platform in the market is mostly based on a single objective collection and collation of tourism information, which only extracts superficial information and has little effect on the development of tourism e-commerce. Therefore, it is necessary to conduct data mining in time and effectively, and find useful knowledge among the massive user data information to serve the strategic decision of enterprise users. Through effective data mining means, various algorithms can be used to identify and screen out the effective data information for tourism e-commerce enterprises, so as to help tourism e-commerce enterprises optimize their e-commerce marketing methods [1].

In view of the above problems, the author of this paper thinks that a tourism e-commerce marketing system can be built by using J2EE technology, JAVA language, Apriori algorithm of association rules and other technologies, which can complete data mining. The basic program developed in this paper, combined with the basic strategy of tourism marketing, provides a technical platform for tourism e-commerce enterprises to combine intelligent application with the realization of tourism marketing needs. Through data analysis, this platform digs out the data of popular scenic spots in different regions and seasons according to different tourists' preferences, and helps tourism enterprises to screen out the most suitable scenic spots and the best priority of scenic spots viewing in planning their travel plans. At the same time, it formulates tourism eating, drinking, playing and buying themes that are more in line with consumers' travel needs, so that consumers can achieve a more convenient, fast, comfortable and free travel experience.

2 Key technology

2.1 Data mining technology

Data mining is the process of extracting hidden information and knowledge that people don't know in advance but are potentially useful from a large number of incomplete, noisy, fuzzy and random data. The task of data mining is to discover patterns from data sets.

Data mining process: Data mining is not our purpose, our purpose is to help the business better, so the first step is to understand the project requirements from a business perspective, and then define the goal of data mining. Try to collect some data, and then explore the data, including data description, data quality verification, etc. This will help you to have a preliminary understanding of the collected data. Start collecting data, clean and integrate the data, and finish the preparation before data mining. Select and apply various data mining models, and optimize them to get better classification results. Evaluate the model, and check every step of building

the model to confirm whether the model has achieved the predetermined business objectives. The role of the model is to find gold mines from data, which is what we call "knowledge". The acquired knowledge needs to be transformed into a way that users can use. The form presented can be a report or a more complex and repeatable data mining process. If the data mining results are part of daily operations, then the follow-up monitoring and maintenance will become important. The data mining flow chart is shown in Figure 1 [3].

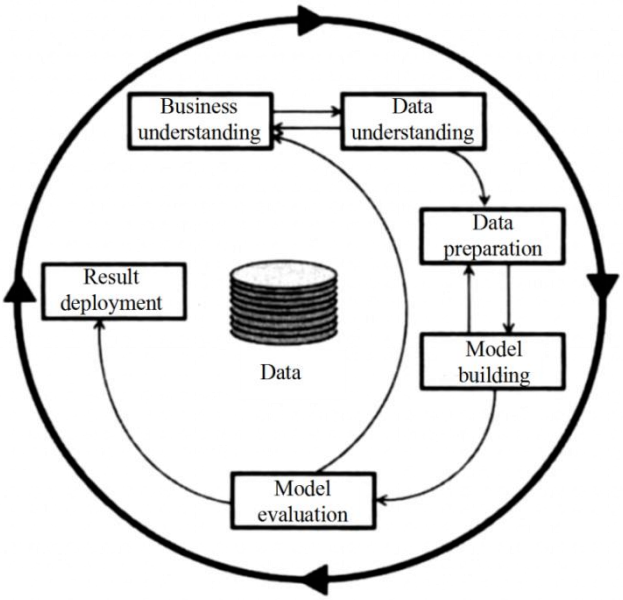

Figure 1: Flow chart of data mining

Data mining method: Data mining is mainly divided into three categories: classification algorithm, clustering algorithm and association rules. These three categories basically cover all the requirements of the current commercial market for algorithms. There are many classical algorithms in these three categories. In December, 2006, the international authoritative academic organization (ICDM) selected ten classic algorithms in the field of data mining, including PageRank of connection analysis, Apriori of association analysis, C4.5 of classification algorithm, Naive Bayes, SVM, KNN, Adaboost, CART, K-Means and EM of clustering algorithm.

2.2 Association analysis Apriori algorithm

Apriori algorithm is used to solve the problem of association analysis of large-scale data sets. Association analysis or association rule learning is to find the implicit relationship between items from large-scale data sets. This relationship can take two forms: frequent itemsets or association rules. Frequent itemsets are collections of items that often appear together, and association rules imply that there may be a strong relationship between the two items. Quantitatively measuring whether a collection of items is frequent or not and quantitatively measuring the relationship between two items use the following concepts: itemsets are collections of items, and items can be commodities, so itemsets are collections of commodities.

Support degree is the proportion of records in the dataset containing the itemset, that is, the frequency of the itemset in the dataset, which is used to measure the frequency of itemsets. Credibility, also known as confidence, is defined for association rules, which indicates the occurrence probability of a certain set under specified conditions, and is used to measure the relationship between items [6].

Looking at the workflow of Apriori algorithm in Figure 2, we can find that Apriori algorithm firstly scans the data set and generates 1-itemset C1 from it. Indicates the Scan data item scanning function, which filters the item set that does not meet the minimum support. Then, the Scan function is called to scan C1 and filter the itemsets that do not meet the minimum support, and the last itemset left is the frequent itemset L1. According to Apriori principle, all supersets of infrequent itemsets are also infrequent, so we don't need to find the combination of these infrequent itemsets. Therefore, in the second iteration, only the frequent itemsets generated in the last iteration need to be newly combined, and then the Scan function is called to check whether the support of the new combination meets the minimum support requirement, and the unsatisfied new combination is filtered. Loop until no new combination can be generated.

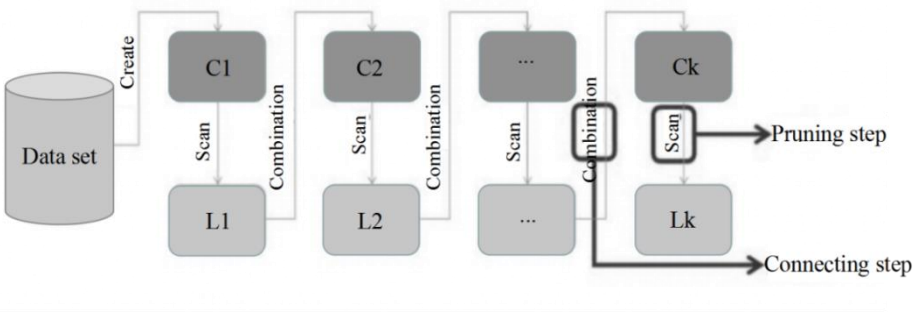

Figure 2: Apriori workflow

2.3 J2EE framework

J2EE is an architecture that uses Java2 platform to simplify the development, deployment and management of enterprise solutions. The foundation of J2EE technology is the standard version of the core Java platform or Java2 platform. J2EE not only consolidates many advantages of the standard version, such as the feature of "write once and run anywhere", JDBCAPI which is convenient to access databases, CORBA technology, and the security mode that can protect data in Internet applications, etc., but also provides full support for EJB (Enterprise Java Beans), Java Servlets API, JSP (Java Server Pages) and XML technology. Its ultimate goal is to become an architecture that can greatly shorten the time to market for enterprise developers. J2EE architecture provides a middle-tier integration framework to meet the requirements of applications that need high availability, high reliability and scalability without too much cost. By providing a unified development platform, J2EE reduces the cost and complexity of developing multi-tier applications. At the same time, it provides strong support for the integration of existing applications, fully supports Enterprise Java Beans, has good wizards to

support the packaging and deployment of applications, adds directory support, strengthens the security mechanism and improves the performance [2].

2.4 Development environment

According to the system development requirements and the use requirements of the above key technologies, complete the configuration and deployment of the development environment. The overall development of the system is based on Windows10.0 operating system, and JDK software needs to be installed to develop Java-based applications. JDK version requires 1.8 or above, and MyEclipse2020 is selected as the Java development environment. The Java development framework is Structs2.3. For the web server, it is very important to support the running environment of the application system. Tomcat9.0 is chosen as the running environment of the system. The Web server is Tomcat8.5, and the database server is MySQL8.0.28.

To connect MySQL with MyEclipse2020, you need to create a new Web project under MyEclipse, and then import the database connection package mysql-connector-java, which is required for development. Then, in window-> open perspective-> myEclipse DatabaseExplorer, right-click the mouse in the left margin of the window that appears-> New, and the databases driver interface appears. Add content: Driver template, select MySQL Connector/J. The URL in connection is changed to jdbc: mysql://localhost:3306/sc. Select the first downloaded driver jar package in Driver JARs. Tructs2 needs to create a new web project when adding in MyEclipse. Right-click Project → MyEclipse → Project Facets → Install Apache Struts (2.3) Facets. Through the introduction of the above key technical theories, we have determined the overall environment of the system development, the configuration of related software and tools, and the technical feasibility of the overall project of the personal income tax management system.

3 Requirements analysis

3.1 Functional requirements

This system uses modular design of functions, which can make the software structure clear and easier to design. According to the characteristics of tourism e-commerce enterprise managers, the user interface of this system is required to be friendly and user-friendly. Tourism e-commerce marketing system studies various data of tourism consumers, so as to help managers and users of tourism e-commerce enterprises to conduct demand analysis from the perspective of consumers' demand for making marketing methods.

Through data analysis, this tourism e-commerce marketing platform digs the data among tourists' favorite tourist attractions for correlation analysis, and guides tourism enterprises to adjust and plan the corresponding tourism plans, thus promoting the marketing innovation of tourism enterprises. The calculation and analysis of consumer's consumption habit data in the process of traveling can help the managers and users of e-commerce enterprises to analyse the consumption preferences of different types of passengers for food, clothing, housing and transportation in the process of traveling. Make tourism e-commerce enterprises help tourists find their own tourism projects faster and arrange and adjust their work and travel plans, so as to achieve a more convenient, fast, comfortable and free travel experience [8].

3.2 Global design

This system adopts multi-layer B/S system architecture, in which each layer completes different functions, so that the tasks and objectives of each layer can be clearly understood when the system is designed, and the development speed and efficiency of the system can be improved. The system design structure is designed according to the structure model of J2EE. The system is developed with Struts2 framework. The system is divided into five layers, as shown in Figure 3, including user interface layer, control layer, business logic layer, data access layer and database layer. Components of the user interface layer are Servlet and JSPs. Servlet and JSP are all used to generate dynamic web pages, accept and check the data input by users, transmit the data to the back-end business logic layer, and return the results processed by the business logic layer to users. The control layer makes the servlet responsible for controlling the jumping function of the client page. The business layer realizes all the business of the system, and the business modules include seasonal analysis, core scenic spot analysis, passenger group structure analysis and tourism habit analysis. These business implementations all need to use Java language for data processing. First, call and read the data of the local e-commerce marketing website of tourism enterprises, then convert the data format into a format that can be compiled by Java and store it in a file, and establish a new database to cache these pre-processed data. After that, Apriori algorithm mining of association rules is carried out, and the statistical comparison results are saved in the system database. Implementation of data visualization by Java Web. JDBC is the technology of connecting database in Java platform, which is the interface standard between Java and database. The logic implementation process needs to access RDBMS through JDBCAPI. In the database layer, MySQL database system is used in this system to store data [10].

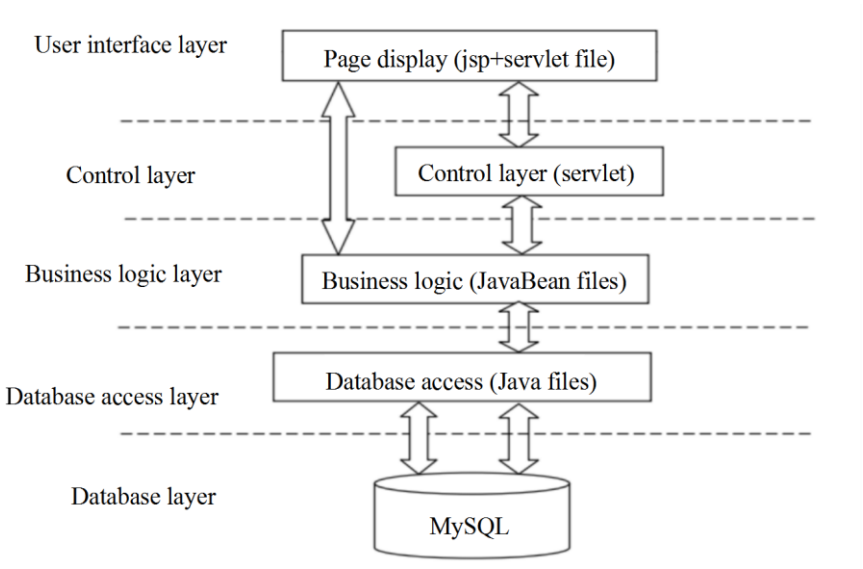

Figure 3: Overall architecture diagram of the system

4 Function implementation

The marketing platform of tourism e-commerce based on big data mining technology is aimed at the users of various tourism e-commerce enterprises. When users log in to the system for the first time, they need to complete the application and registration of account according to the operation guide. After approval, users can log in with a unique username and password. For the first time, the logged-in user needs to connect to the e-commerce marketing website of the enterprise to obtain the access and calling right of the local database. After that, users can access the four main functional modules of the system, including seasonal analysis, core scenic spot analysis, tourist group structure and tourism habit analysis. The analysis function of each module of the system is mainly developed in Java language according to Apriori algorithm of association rules in data mining technology [7].

Click on the seasonal analysis function, and the system can analyse the seasonal peak flow of tourists participating in scenic spots in various regions according to the data comparison of peak passenger numbers in various seasons and months. Take Hebei Province as an example. When visiting Hebei Province in summer, most tourists who choose Chengde Mountain Resort scenic spots will also be interested in Aranya Church scenic spot in Qinhuangdao to make the next arrangement and choice. Therefore, the managers and users of tourism e-commerce enterprises can arrange the Chengde Mountain Resort and Aranya Church scenic spot in the same travel planning itinerary for bundled sales in summer according to the data analysis results. It can help enterprises optimize the customized itinerary of passengers for different seasons. Pay attention to the staged cooperation with offline travel agencies, deepen the products to specific cities, close the distance with consumers and increase the market share. Click to enter the core scenic spot analysis, and you can view and compare the traffic data of different scenic spots in each region. Taking Xi 'an as an example, when many customers visit one of the scenic spots in Sajin Bridge and Muslim Snack Street, their interest in visiting the other scenic spot is relatively decreased, so enterprise users should try their best to avoid arranging the two scenic spots in the same travel plan when making travel plans. Part of the data comparison can help users to separate out the best priority of scenic spot planning and the priority of scenic spot publicity and marketing when making tourist plans [4]. In cooperation with key tourist attractions, online travel enterprises can provide tourist markets for tourist attractions, with a view to deepening cooperation, grasping the target market and expanding market share. Click on the analysis of travel habits, and you can see the comparison of spending power and spending frequency of different types of travellers, such as family trips, couples trips, young individual trips, and elderly group travellers, in terms of food, accommodation, transportation, travel, shopping, and entertainment. Thereby helping e-commerce enterprise managers and users to analyse the consumption preferences of different types of passengers for food, clothing, housing and transportation stalls during the travel process. For example, family travellers are more inclined to consume overall housing comfort, and elderly groups lack interest in shopping. This function is beneficial for enterprise users to target different consumer groups of tourists, and tourism websites should actively cater to them and work out tourism schemes that are in line with tourists' personal interests.

5 Conclusions

It is important and meaningful to write data mining algorithms to solve data problems in the network environment. Using data mining technology to realize the tourism e-commerce marketing system is a new research idea, which not only carries out experiments on data storage and data calculation, but also enhances the practicability of the tourism e-commerce marketing system and the user experience of the system. This paper presents a tourism e-commerce marketing system based on data mining technology, which verifies the feasibility of processing tourism data, but there are still some aspects to be further studied. The data source of this paper is relatively single, so we need to collect more abundant data in the future research work, consider the problem comprehensively and avoid contingency and one-sidedness. The marketing aspect of this paper is relatively broad, but for a large amount of data, it has not done more detailed and specific processing to improve the effect of helping marketing.

References

[1] Chen Qiuyang (2014). Design and Implementation of Precision Marketing System Based on Data Mining Technology. Zhejiang University.01.

[2] Chen Sha (2019). Research on Marketing Strategy of Tourism E-commerce Enterprises Based on We Media. Marketing Exploration (Information Edition).11.

[3] Guo Yourong (2016). Marketing Strategy of Personalized Tourism Products of Tourism E-commerce under "internet plus"—Taking Tuniu.com as an Example. Reformation & Strategy.02.

[4] Huang Chaoran (2013). Design and Implementation of Precision Marketing System Based on Data Mining. South China University of Technology.05.

[5] Li Jun (2016). Research on Agent-based Distributed Marketing System of Chinese Tourism Destinations. Beijing Technology and Business University.06.

[6] Ma Honghong (2019). Comparative Analysis of China's Tourism E-commerce Marketing Models —Taking Ctrip, eLong and Qunar.com as Examples. China Collective Economy.10.

[7] Peng Zhiyong, Long Hu (2020). Application of Machine Learning Algorithm in Mining Tourism Data. Tourism Today.02.

[8] Ren Jin, Li Xueyuan, Sheng Chen (2020). Research on Attachment Behavior of Tourism E-commerce Website Based on Goal Framework Theory. Journal of Taishan University.04.

[9] Xi Xi (2019). Marketing Strategy Analysis of Tourism E-commerce Platform in China.SP.05.

[10] Yin Fangchao (2020). Design and Application of Hebei Culture and Tourism Platform Based on Big Data Technology. Digital Technology and Application. 01.