Establishing an intelligent and smart tourism model using data mining in the context of big data

Xiaoxue Jia (Sanbo0302@126.com)
Zhengzhou Tourism Vocational College

Research Article

Keywords: Big data, data mining, artificial intelligence, smart tourism, intelligent tourism model

Posted Date: August 8th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3073900/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Establishing an intelligent and smart tourism model using data mining in the context of big data

Xiaoxue Jia

Finance Office, Zhengzhou Tourism College, Zhengzhou, Henan, 450000, China
Email: Sanbo0302@126.com

Abstract: Nowadays, with the rapid development of information technology (IT), the tourism industry has begun to apply it in the tourism sector for the establishment of a smart tourism model. Smart tourism is formed based on the expansion of traditional tourism, focusing on personalized user experiences through TIYA, which is a social networking application. Its core is IT, such as cloud computing, the Internet of Things (IoT), artificial intelligence (AI), and big data technology. Smart tourism is a complete tourism service system that improves the convenience of tourists. Keeping in mind the vital role of smart tourism in developing the economy of a country, this study establishes a smart tourism management model based on data mining in the context of big data, with the purpose of strengthening the accuracy of smart tourism management and improving customer satisfaction with tourism. Before constructing the intelligent tourism management model, this study first describes the data mining technology in detail, introduces the Naive Bayes (NB) algorithm and the improved algorithm based on Apriori, analyzes the characteristics of tourism data, and establishes an intelligent tourism management model on this basis. In the proposed model, the government sector, the tourism industries, the community residents, tourists, and other forces are brought into full play, and a complete intelligent tourism management model is established. Using the proposed model, all the stakeholders can improve their own value and meet their tourism needs. Taking Beijing as an example for the experimental work, the improved algorithm based on Apriori is used to mine the tourist interest points. After evaluating the proposed model, it is concluded that the most interest points are obtained when the bandwidth parameter is set to 0.02. After the clustering operation, the first, second, and third tourist attractions in Beijing are obtained, and the results are consistent with the tourist interest. At the same time, the time index of three different mining algorithms (improved MapReduce algorithm (MA), the MapReduce algorithm (MRA), and the k-order MapReduce parallel algorithm (MRKA) based on the improved Apriori algorithm) under different minimum supports is compared. The results show that the MA algorithm can mine the required information in the shortest time.

Keywords: Big data, data mining, artificial intelligence, smart tourism, intelligent tourism model

1. Introduction

The travel of people between different geographic locations is referred to as tourism and is also known as a service activity. In other words, travel for pleasure, business, or commercial reasons to locations outside of one's local area is referred to as tourism, which is a cultural, social, and economic phenomenon [1]. In terms of business and marketing, tourism is crucial to the political, social, intellectual, and economic growth of the majority of nations [2]. It is very interesting to note that the tourism industry is expanding so quickly, which we can observe from the fact that there were a total of 546 million visitors in 2006 across the globe, and this number reached 1.34 billion worldwide in 2016. By 2030, the number of tourists across the globe is anticipated to increase even
further, reaching 1.9 billion. By providing millions of jobs globally, such statistics give significant insights into the influence and impact of tourism on humanity and society [3]. The tourism industry makes it possible for people to comprehend the whole development of travel from the perspectives of providers and consumers [4]. Although it has always been a feature of human activity, tourism is a phenomenon that is becoming more and more significant on the social and economic fronts [5].

Tourism has experienced a boom in recent years and has become a significant economic industry not just in the European Union (EU), but also in other regions of the world. In 2017, the EU accounted for around 510 million, or 42.6 percent, of all international visitor arrivals worldwide [6]. Additionally, recent research in several nations has shown the significance of tourism as a source of economic development [7–9]. With the improvement of the living standard, more and more people like to relax and improve their quality of life through tourism. The rapid development of the tourism industry has made all scenic spots very popular [10]. During the season of tourism, due to the large number of tourists, it is difficult for the tourists to arrange proper accommodation and to visit famous tourist spots with ease. As a result, the tourists lack basic tourism resources and fail to arrange their personal travel plans according to the resources. Improving the travel experience and comfort results in satisfaction in the process of tourism [11]. The tourist sector has been significantly impacted by the transformation and revolution of the Internet and IT [12]. Tourists are increasingly using their smartphones to help with navigation and site discovery in smart tourism [13]. To ensure that tourism is a win-win activity for everyone, the tourism sector has to work with communities and individuals. Hence, significant improvements to tourist development and management approaches are always needed [14].

The best tours can be planned with the help of tourism management systems, which also give system users access to pertinent information like weather, location, geography, and festivals. A tourist management system has a significant influence on e-tourism, which may also be used in industrial and personal trips [15]. All travel sector processes are digitalized under the e-tourism system. E-tourism also strives to improve the effectiveness of communication, trade, and payment systems [16]. Over the last several years, the terminology "e-tourism system" has quickly grown to denote the ICT-related sectors and services that are heavily reliant on and adopted by the general public. The tourist industry has been recognized as one of the leading industries that have rapidly changed with e-tourism and offer "smart tourism" as part of its expansion technology. According to Buhalis et al. [13], the key goals of the smart tourism industry are to deliver actual data, situational awareness, and personalization. This procedure improves the tourists' quality of life and provides a practical and engaging atmosphere to users of the system [17].

The tourism sector generates an enormous volume of data, which is not possible for traditional database systems to store and analyze. This is a serious problem that needs proper attention and consideration. Keeping the vital role of big data and data mining techniques in various sectors in consideration, this study proposes an intelligent tourism management model based on data mining techniques in the context of big data [18–19]. The proposed intelligent tourism management system combines the three aspects of tourism, i.e., tourism management, tourism service, and tourism marketing, to obtain and analyze control data on tourism public opinion. Further, it digs deeper into the points of interest of tourists and hot spots of tourism, guides the creation of marketing themes for tourism products, and establishes a complete model of intelligent tourism management. This paper uses the latest data mining technology to analyze tourist resources, the most interesting tourist attractions, and other data to analyze the preferences of tourists.
The main innovations and contributions of this study are listed as follows:

- This study focuses on the implementation and analysis of data mining algorithms in the tourism sector. The utilized data mining algorithms include the naive Bayesian and the improved Apriori algorithms.
- The utilized data mining algorithms obtain the data characteristics of the tourism industry for analysis, based on which a data mapping model using tourism knowledge is established. Furthermore, the characteristics of the traditional and the proposed intelligent tourism management models are compared, and on this basis, a large-data tourism management model is developed that controls the tourism sector.
- In this study, the performance of the improved MapReduce algorithm (MA), MapReduce algorithm (MRA), and k-order MapReduce parallel algorithm (MRKA) based on the improved Apriori algorithm is compared. In addition, the minimum degree of support is selected in this process, and the running cycles obtained by the three algorithms at different minimum scales are compared. The simulation and analysis results prove the significance of the proposed model.

The remaining paper is organized in the following order: Section 2 shows the related work accomplished in the domain of the tourism sector. Section 3 explains the data mining techniques used in this study. Section 4 illustrates the intelligent travel management model based on the concept of data discovery. Section 5 represents the analysis and simulations of the intelligent travel management model based on data discovery. Finally, the overall theme, findings, and concept of the smart tourism management system are concluded in Section 6.

2. Related Work

Tourism is a cultural, social, and economic phenomenon. Therefore, by traveling to other nations or regions, a large number of individuals worldwide may significantly affect their economies [20]. Generally, tourists are anxious to see several areas of interest, but owing to time and budgetary restrictions, it is not possible to visit every location in which they are interested. As a result, the tourists are forced to pick only a few locations from the draft of the selected locations [21–22]. Therefore, while organizing a trip, travelers must decide which route among the daily routes is the most interesting and feasible one, depending on their time and financial constraints [22]. Nowadays, smart tourism is in a fast-developing state, and each country has its own rules and policies about tourism. The development of the smart tourism industry cannot be separated from advanced technology [23]. Lin F. N. et al. [24] stated that smart travel is a mobile travel information service platform developed on the basis of a smartphone platform. According to their study, tourists can download software on their mobile phones to get information about nearby scenic spots and get coupons. Similarly, Brindha et al. [25] have analyzed intelligent travel from the perspective of tourism service quality and established a service quality satisfaction index of intelligent travel scenic spots based on the SEV QUAL difference model. Fernandes et al. [26] have used the MD-Apriori algorithm to design an intelligent recommendation system for travel services that can be used to mine the generated rules and then recommend the analysis results to users.

Similarly, Yochum et al. [27] developed a constraints-based travel recommendation system that combines user interaction and conversation modes to use an incremental approach for locating locations of interest to users. Zhou X et al. [28] developed a distributed intelligent recommendation system that improves retrieval convenience, but at the same time, it is not an intelligent travel service.
recommendation system. Another researcher established a management platform and perception system to analyze intelligent travel, so as to study the established management platform and display the intelligent travel management platform through the model [29]. Beraldi et al. [30] pointed out that smart travel can be extended to other areas, such as green, clean, quality, and ethics. Another scholar, Kumar, defined smart tourism as a valuable and intelligent model of the relationship between cities and tourists using mobile technology, pointing out that smart tourism is not a consumption model but a form of tourist participation based on data [31]. Arif et al. [32] stated that smart travel is a combination of cloud computing, the Internet of Things (IoT), and other technologies, using portable smart terminals to access the mobile internet and the Internet in order to use and intelligently perceive a variety of travel information [33]. Muniz et al. [34] implemented intelligent management of the tourist experience by integrating destination management organizations and user knowledge management and comprehensively analyzing CKM models and frameworks to improve and innovate the tourist experience and tourist products.

The ability of tourist firms and places to not only gather massive amounts of data but also to effectively store, interpret, integrate, evaluate, and apply big data to create tourism processes, services, and entrepreneurial ventures clearly forms the foundation of the idea of smart tourism. The technical underpinnings of smart tourism are multifaceted and include pervasive architecture, portable and context-aware information management, and increasingly multidimensional connectivity that supports interconnections with the traveler's immediate environment as well as the larger community and society in general. The foundation of smart, sustainable tourism is the gathering, exchanging, and processing of data produced in many parts of the system, affecting the customer, the enterprise, and the location overall. Big data analytics is undoubtedly a toolbox with a wide range of data types, analytical techniques, and commercial applications [35]. Big data analytics enhances our ability to comprehend the consumer market at a degree, range, and depth that are unheard of when compared to conventional research and innovation approaches [36]. This study focuses on the implementation and analysis of data mining algorithms in the tourism sector. The utilized data mining algorithms include the naive Bayesian and the improved Apriori algorithms. In addition, the data mining algorithms used in this study analyze the characteristics of tourism data, and a data mapping model is established using the tourism data. Furthermore, the characteristics of the traditional and the proposed intelligent tourism management models are compared, and on this basis, a large data tourism management model is developed that controls the tourism sector. It is anticipated that the model developed in this study will be of great help for the tourism industry and the tourists as well.

3. Overview of Data Mining Algorithms

3.1 Definition and process of data mining

Data mining (DM) represents the process of finding new, efficient, and valuable data within a large amount of data while being understood by users. Some people describe the term data mining as the basic process of mining and finding database knowledge. Here are some of the basic steps for data mining.

- **Data cleanup**: In this step, the noise is removed from the original data, or processed inconsistent data is removed from the original data. Due to the lack of uniform data and a large difference between structures, data redundancy, incomplete information, etc., some data is lost and incomplete, so that data needs to be repaired or removed from the original data.

- **Data integration**: In this step, various data sources are combined as a whole into a single source.
This indeed is a vital step for further processing.

- **Data selection**: In this step, only problem-related data is selected from a large amount of data. Extracting problem-related data from the database and analyzing various related data will ignore the data that is of no value.

- **Data transformation**: This step transforms the original data in the database and unifying all kinds of original data to form ideal mining that simplifies the difficulty of data mining, speeds up data mining, and reduce hardware investment.

- **Data mining methods**: This paper uses different data mining methods to complete data mining operations and uses mature mining tools to design mining methods that meet the requirements.

- **Model evaluation**: Usually, the results of data mining models are unexpected, and some of them can reflect the implicit relationship between data. Further, its results depend on the nature of the data, sometimes an algorithm performed better on one kind of data but performs poorly on the other type of data. Therefore, the evaluation of an algorithm is also of great significance. For this purpose, different evaluation metrics are used such as accuracy, precision, recall, etc.

- **Knowledge representation**: Using visualization technology or other knowledge representation technology to acquire data mining knowledge, users can intuitively view the obtained mining data and can have a deep understanding of mining knowledge.

Figure 1 shows the completed data mining process.

![Figure 1. An overview of complete data mining process architecture](image-url)

Each process in data mining is related to one another. Mining is based on the initial amount of mining results or interaction between different users, and feedback is added to the mining conditions to mine more valuable data. Therefore, data mining needs many iterations, corrections, and loops to implement [37].

### 3.2 Naive Bayes (NB) algorithm

NB is a probabilistic data mining algorithm, used for a variety of classification and categorization tasks. Bayes theorem is the foundation for the NB classification algorithm. NB algorithm is used to calculate different types of probability under different conditions. The high
probability of a certain category indicates that the item-instance/subject belongs to that particular category. NB needs to be trained on some data before using it for the classification or categorization tasks. To classify a problem based on the Bayesian theorem, it is necessary to first determine the characteristic of the sample and determine the probability of the sample, which is calculated by the following formula:

\[ p(y|x) = \frac{p(x|y)p(y)}{p(x)} \quad (1) \]

By expanding eq. (1), we obtain:

\[ p(y = c_k | x) = \frac{\prod_{i=1}^{M} p(x'|y = c_k) p(y = c_k)}{\sum_{k'} p(y = c_{k'}) \prod_{i=1}^{M} p(x'|y = c_{k'})} \quad (2) \]

In eq. (2) \( x \) is the eigenvector and \( M \) is the dimension of \( x \). In this study, the simple Bayesian model is used to manage the tourist routes. For the routes in which the tourists are interested, the NB finds the keywords and then filters out the other routes as well that meet the requirements of the tourists. At the same time, the recommended travel routes meet the personal selection and the needs of tourists.

### 3.3 Improved algorithm based on Apriori

The frequently encountered item sets and relevant association rules are mined using a method called the Apriori algorithm. The Apriori algorithm frequently operates on a database that has a large number of transactions. But at the same time, the conventional Apriori algorithm has some shortcomings. Based on the nature of association rules, this paper re-improves the Apriori algorithm and presents an improved Apriori algorithm in order to apply it in the smart tourism sector. This algorithm uses the concept of vertical distribution to convert transaction data on the D database to post-matrix statistics.

**Step 1:** Convert the following D matrix database:

\[
D = \begin{bmatrix}
  d_{11} & d_{12} & \ldots & d_{1n} \\
  d_{21} & d_{22} & \ldots & d_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  d_{m1} & d_{m2} & \ldots & d_{mn}
\end{bmatrix} I_1
\]

\[ dij = \begin{cases} 
0 & I_i \notin T_j \\
1 & I_i \in T_j
\end{cases} \quad (4) \]

Here \( T_i \) represents the transaction record data on D database, \( I_1 \) is the project data, whereas the \( I_i \) project support is calculated from the following formula:

\[ \text{sup port}(I_i) = \sum_{j=1}^{\text{sup}} d_{ij} \quad (5) \]

**Step 2:** Define the \( I_i \) corresponding to each project:
Step 3: The vector product of the \( \{I_i, I_j\} \) item set is given as follows:

\[
D_i = \begin{bmatrix}
  d_{i1} \\
  d_{i2} \\
  \vdots \\
  d_{in}
\end{bmatrix}
\]

\[
d_{ij} = \begin{cases}
  0 & I_i \notin T_j \\
  1 & I_i \in T_j
\end{cases}
\]

(6)

Where the expression \( \wedge \) denotes logic and operation, while the following formula is used to calculate the \( \{I_i, I_j\} \) item-set support:

\[
\sup \port \{I_i, I_j\} = \sum_{r=1}^{m} (d_{i1} \wedge d_{j1})
\]

(7)

Step 4: The \( \{I_1, I_2, \ldots, I_k\} \) item set vector product is defined by the following formula:

\[
D_{1,2,\ldots,k} = D_1 \wedge D_2 \wedge \ldots \wedge D_k
\]

(8)

The following formula is used to calculate the support count for \( k \) item-set \( \{I_1, I_2, \ldots, I_k\} \):

\[
\sup \port \{I_1, I_2, \ldots, I_k\} = \sum_{r=1}^{m} (d_{i1} \wedge d_{i2} \wedge \ldots \wedge d_{ik})
\]

(9)

4. Intelligent travel management model based on data discovery

4.1 The analysis of tourism data

This paper chooses the tourism information from 2019 to 2020 using the relational database of a tourism company in Shaanxi province. The characteristics of tourism data include the feature tables of tourist attractions, the information tables of tourists, and the travel record tables. Based on the characteristics of tourist attractions and the tourists, this paper establishes a data mapping model based on tourism knowledge, as shown in figure 2 [38].
Two variables determine how tourists have the awareness of specific tourist sights. The first is the tourists' own academic background, level of energy, accumulated knowledge, etc. Secondly, media advertising, others' recommendations, web searches, etc., are a part of external factors. Internal factors are the accumulation of individual tourists over a long period of time, which is usually difficult to change. Since external factors are influenced by internal factors, we can identify the best external factors from the fundamental characteristics of tourists, such as hobbies, gender, income, and age, to alter the level of interest in the tourist attractions and increase tourists' satisfaction.

4.2 Intelligent travel management model

This paper proposes an intelligent travel management model based on the large data platform, which can provide preferential ticket information, travel route quotations, etc. to tourists and give travel suggestions, guidelines, and routes based on this data. The proposed intelligent tourism management model is based on a big data platform that integrates various tourism information to provide a personalized service to tourists. The main items of this service include pre-purchase tickets, travel routes, hotel tickets, visa services, taxi call services, travel insurance, travel blogs, recording travel information, etc. It can give real-time feedback to tourists and send information such as accommodation, transportation, catering, the intensity of tourists, and the direction of tourism spots in areas where tourists are located. According to this information, tourists adjust their personal travel plans and travel routes, as shown in figure 3.
By establishing an intelligent tourism management model, we can better deal with the problems existing in tourist areas, community residents, tourism enterprises, and governments during the peak period of tourism, realize the sharing and communication of interests among individuals in tourism activities, and better manage regional activities, operation modes, etc. [39]. In-depth analysis of the mechanism of the intelligent management model, and according to four different stakeholders, the smart tourism management platform is divided into traveler sub-platform, tourism department platform, tourism enterprise platform, and community resident platform. The main differences between the platforms are reflected in the groups they are facing. The common point is achieved by the data on tourism activities.

**4.3 Intelligent travel market expansion strategy in the background of big data**

Compared with traditional tourism methods, the tourism industry has undergone significant changes in all aspects since it entered the era of big data. People use advanced technology to obtain tourism information, and tremendous changes have taken place in tourists' demand, travel modes, service types, and marketing modes. Table 1 shows a comparison between traditional travel modes and the travel modes using the concept of big data.

<table>
<thead>
<tr>
<th>Traditional tourism mode</th>
<th>Category</th>
<th>Tourism in the era of big data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass promotion of</td>
<td>Ways to obtain tourist</td>
<td>Accurate promotion to improve</td>
</tr>
<tr>
<td>information, investment</td>
<td>resources</td>
<td>referral rate and revisit rate</td>
</tr>
<tr>
<td>and competition for tourists</td>
<td></td>
<td></td>
</tr>
<tr>
<td>People with tourism needs</td>
<td>Tourist type</td>
<td>Private customization,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>personalized demand and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>precision marketing</td>
</tr>
<tr>
<td>Group tour and sightseeing</td>
<td>Tourist demand</td>
<td>Deep experience and theme</td>
</tr>
<tr>
<td>tour</td>
<td>product type</td>
<td>formulation</td>
</tr>
<tr>
<td>Homogeneous products</td>
<td></td>
<td>Featured products integrating</td>
</tr>
<tr>
<td></td>
<td></td>
<td>leisure experience and popular</td>
</tr>
<tr>
<td></td>
<td></td>
<td>science education</td>
</tr>
</tbody>
</table>
In the upstream position of the intelligent travel management mode, the travel operators and tourism products should focus on analyzing and mastering the attributes of various tourism products, digging deeply into the core content and attractiveness of the products themselves, combining innovative products, optimizing product quality and service, and packaging the special products in the scenic spots, which will become the main way for the travel enterprises to expand the market [40]. The downstream of the intelligent travel management model are the tourists. Enterprises should dig deeply into the tourists' travel needs, send them products that meet their attributes, take measures to dig potential users, and stimulate tourists' purchasing behavior, which is also the key to expanding the market for managers of tourism enterprises. Tourism marketing plays a connecting role in the model of exhibition tourism management, which can better speed up the communication between the internal and external enterprises, and realize the further expansion of the tourism industry. Figure 3 shows the strategy of tourism market expansion under the background of big data.

Based on the background of big data, optimize and improve the market expansion strategy, and penetrate all kinds of data in each process, which can change the traditional one-way, single transmission path to feedback, two-way process. The intelligent travel management model downstream travelers experience and introduces tourism products and services, further optimize the products using evaluation data, and fundamentally adjust and improves the tourism market development strategy. At the same time, the large data platform is used to ensure that the marketing mode of tourism products, tourists, and enterprise expansion strategies are all in a big data environment, and to analyze a variety of forward and reverse data feedback mechanisms that can greatly improve the utilization rate and value of large data. Based on the concept of big data, data mining algorithms, and the development status of the tourism market, this paper achieves the expansion of the tourism market from four points: optimizing product services, locating regional markets, analyzing tourists' audience, and innovative marketing means, keeping the intelligent tourism industry in a benign competitive state, enabling tourists to experience a higher quality of tourism, and enterprises to obtain a higher profit space in the process of tourists' travel.
5. Analysis of intelligent travel management model based on data discovery
5.1 Analysis of tourist interest point mining based on improved Apriori algorithm

This paper builds an intelligent travel management model based on the concept of both data mining algorithms and big data. In order to test the effectiveness of the developed model, the crawler technology is used to obtain the Geotag and context information of Beijing on the Flickr website as the experimental data. The main goal of experimental work and analysis is to dig more parameters to meet the requirements of tourists' points of interest. In the experimental work, we set the bandwidth parameter to 0.02. The experimental results attained using the improved Apriori algorithm are shown in figure 5.

![Figure 5. Mining results of tourists' points of interest using improved Apriori algorithm](image)

The results shown in figure 5 are achieved using the bandwidth parameter setting of 0.02 and the number of points of interest obtained under this parameter setting is 709, which delineates the different points of interest of tourists, and the authenticity of tourist’s activities is strong. A large number of points of interest will prolong the running cycle of the algorithm and make it more difficult to obtain tourist routes of scenic spots. Further, we observed and analyzed the distribution of points of interest under this parameter setting, using a mean-shift clustering operation with a bandwidth of 0.02. Based on this parameter setting, 709 tourists' points of interest in Beijing are obtained. According to the results of mining the points of interest of tourists in Beijing, the information of scenic spots in different districts of Beijing can be obtained by statistical analysis of photo quantity. The results of mining the popular scenic spots in Beijing are listed in Table 2.

| Table 2. Excavation of popular attractions results in Beijing |
According to table 2, the most popular attraction areas are Shichahai, the imperial palace, the summer palace, the Olympic park, and the Badaling great wall. According to the mining data, these scenic spots match the attractions of interest of tourists and are also becoming the favorite tourism spots for the tourists of other regions. This shows that the data mining results meet the real needs of tourists, provides better recommendations for tourists, and provide more choices for tourists under the smart travel management model.

5.2 Analysis of tourist activity model

Based on the data of the points of interest mined above, we can get the tourists' patterns of interest and other activities that they like the most. For this purpose, the association rules are used to mine the initial dataset, which contains 101,565 transactions. As the number of transactions is above 1 lac, which increases with each passing day and is almost impossible to process it with the help of traditional data processing techniques. Here comes the role of big data, which easily processes and analyzes billions of instructions in a very short amount of time. In this paper, the performance of the improved MapReduce algorithm (MA), MapReduce algorithm (MRA), and k-order MapReduce parallel algorithm (MRKA) based on the improved Apriori algorithm are compared. Further, the minimum degree of support is selected in this process, and the running cycles obtained by the three algorithms at different minimum scales are compared. The running time of three different mining algorithms is illustrated in figure 6.

![The time index of three mining methods with different minimum scale](image)

Figure 6. Time index of the three big data mining algorithms at various minimum scales

Figure 6 shows that when the number of candidate sets increases and the minimum support...
level decreases, the corresponding three algorithms run longer and takes more time in processing the transactions. Among the three mining algorithms, the MRKA algorithm takes more time and is considered worse in terms of processing time. The basic logic of taking more time for processing is that this algorithm needs MapReduce to be implemented multiple times when it is applied, while other algorithms can be completed in only two steps, and a large number of joint records are produced as an output during production. At the same time, there is a big difference between the running cycle of the MRA and the MRKA algorithm when the minimum scale is reduced. Because the MRA algorithm produces a large number of local frequent itemsets in the first MapReduce process, adding two MapReduce processing cycles and directly extending the running cycle in the case of a large number of candidate sets. In addition, the improved MapReduce algorithm based on Apriori (MA) used in this paper has the shortest running time as compared to the other algorithms.

The size of the dataset grows as a result of the ongoing expansion of transactions. Therefore, it would be useful to assess how well the algorithms scale, especially when dealing with larger datasets. Figure 7 depicts our comparison of the MRA, MRKA, and enhanced Apriori MapReduce Algorithm (MA) in terms of scalability performance. In our case, the dataset ranges in size between 100,000 and 500,000.

![Figure 7. Scalability of MA, MRA, and MRKA algorithms](image)

Figure 7 shows that the scalability performance of our proposed algorithm (MA algorithm) is better in terms of execution time as compared to other algorithms (MRA and MRKA). MA gives better scalability by processing the transaction more efficiently as the dataset size increases.

We compared the overall performance of the MRA, MRKA, and MA algorithms in terms of different performance measures by examining the F1-score, recall, and precision as shown in Figure 8 in order to assess the quality of the results that each algorithm produced. Figure 8 demonstrates that the MA algorithm outperforms MRA and MRKA in terms of F1 score, recall, and precision. It denotes that, in comparison to other algorithms, the patterns or relationships found by MA are more precise or of higher quality.
In this paper, an improved algorithm based on Apriori (MA) is used to obtain the frequent activity model of tourists. The results obtained by mining data from one frequent set in the activity model are shown in table 3. The results obtained via the improved algorithm based on Apriori are shown in figure 9.

**Table 3. Frequent item set mining results for visitor activity modes**

<table>
<thead>
<tr>
<th>Frequent Item set</th>
<th>Degree of support (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperial Palace</td>
<td>38%</td>
</tr>
<tr>
<td>Summer Palace</td>
<td>19%</td>
</tr>
<tr>
<td>Tiananmen</td>
<td>18%</td>
</tr>
<tr>
<td>Badaling Great Wall</td>
<td>14%</td>
</tr>
<tr>
<td>Temple of Heaven</td>
<td>11%</td>
</tr>
</tbody>
</table>

Among the mining results of one frequent item set, as shown in Table 3, the highest degree of support was observed for the Imperial Palace, i.e., 38%. The second highest degree of support was observed for the Summer Palace, i.e., 19%, and Tiananmen attained the third highest degree of support, 18%. A low degree of support of 11% was observed for the Temple of Heaven, as shown in table 3 and figure 9.
Entering the era of big data, the scale of the internet, mobile networks, and social networks continues to increase, leading to an exponential increase in the amount of generated data. Nowadays, we are living in an era of data mining, artificial intelligence (AI), deep learning (DL), and big data techniques, which are applied extensively in different fields of life. Big data and data mining have various applications in different sectors, such as healthcare, industries, sports, tourism, etc. Recently, the tourism industry has used advanced data mining methods to change the travel mode, providing personalized travel modes to tourists, making a huge change in traditional travel modes, and allowing people to experience the change of life with intelligent technology. Therefore, this paper uses data mining technology against the background of big data to build an intelligent tourism management model, provide high-value tourism services and products to tourists to meet their basic needs, and fundamentally improve the management ability and service quality of tourism enterprises, which has become the main driving force for the development of China's tourism industry. Taking Beijing as an example, the improved Apriori algorithm is used to mine the points of interest of tourists going to Beijing. Based on the mining data analysis, the most points of interest are obtained when the bandwidth parameter is 0.02. Comparing the time indexes of the three mining algorithms with the minimum support degree shows that the algorithm takes the shortest time.

6. Conclusions

Figures 9. Results obtained via the improved Apriori (MA) algorithm

Funding: This research work does not receive any funding.
Conflict of Interests: Weizeng Huang declares that he has no conflict of interest.
Ethical approval: This article does not contain any studies with human participants or animals performed by the author.
Availability of data and materials: Data are available on reasonable request from the corresponding author.

References
[1]. A. Kontogianni, K. Kabassi, and E. Alepis, "Designing a smart tourism mobile application: User modelling through social networks’ user implicit data," in International Conference on Social
Informatics, 2018, pp. 148-158.

[19]. V. Ndou and G. Mele, "Social big data for sustainable smart tourism development: a conceptual
[34] Y. c. Shi and U.-K. Lee, "The Impact of Restaurant Recommendation Information and Recommendation Agent in the Tourism Website on the Satisfaction, Continuous Usage, and
[35]. Z. Xiang, D. Wang, J. T. O'Leary, and D. R. Fesenmaier, "Adapting to the internet: trends in
[36]. Z. Xiang and D. R. Fesenmaier, "Big data analytics, tourism design and smart tourism," in
Personalized Route Recommendation System for Constraint Tourism and Travel," 2021.
for mobile based group and personal decision support system," Register: Jurnal Ilmiah Teknologi