

## Research Article

# Tourist Attraction Recommendation Method Based on Megadata and Artificial Intelligence Algorithm

Laiyan Yun,<sup>1</sup> Huihua Jiao<sup>2</sup> and Kai Lu<sup>3</sup>

<sup>1</sup>College of Finance and Economics, Hainan Vocational University of Science and Technology, Haikou, 571126 Hainan, China

<sup>2</sup>Inetwork and Educational Technology Center, Qiongtai Normal University, Haikou, 571127 Hainan, China

<sup>3</sup>Department of Public Safety Technology, Hainan Vocational College of Political Science and Law, Haikou, 570100 Hainan, China

Correspondence should be addressed to Huihua Jiao; jhh@mail.qtnu.edu.cn

Received 21 May 2022; Revised 6 July 2022; Accepted 15 July 2022; Published 3 August 2022

Academic Editor: Jun Ye

Copyright © 2022 Laiyan Yun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As China's economy continues to grow of informational technology and mobile Internet industry, the online tourism industry has received more and more extensive attention and use. However, as an emerging industry, users often need to spend a lot of time to choose travel services that match their needs because of the complex amount of relevant information. Under such circumstances, this paper studied the recommendation method in travel platform. First, the big data is used to extract user data. Secondly, the current online travel business recommendation for users has the problem of low accuracy. The reason is that the services provided are still in traditional recommendation algorithm. In this paper, the Bayesian network is used to evaluate the user's attribute preference and generate a data model, using effective methods in artificial intelligence algorithms to improve collaborative filtering algorithms and finally generate hybrid recommendation algorithms. Compared with the traditional recommendation method, the experimental results showed that the research can improve the recommendation accuracy of tourist attractions by 6.55%, increase the user's satisfaction for the platform, and enhance the visit rate and retention rate of the tourist attraction recommendation platform.

## 1. Introduction

At present, in order to accurately grasp the user's preferred tourist attractions and improve the visit rate and retention rate of the tourism recommendation platform, the tourism industry has gradually deepened the research on the recommendation algorithm for tourist attractions. However, the current personalized recommendation algorithm still has the problem of low recommendation accuracy. This paper applies big data and artificial intelligence algorithms to improve the recommendation method, which is conducive to improving users' satisfaction with the platform and improving the visit rate and retention rate of the tourist attraction recommendation platform. Improving the visit rate and retention rate of the tourist attraction recommendation platform, it is of great significance to the good development of the online tourism industry.

Liu et al. proposed that it is the most widely used of the personalized recommendation methods. The principle of the

algorithm is based on the user's model labels and the user's ratings and comments on the recommended entities [1]. Han et al. subdivided the collaborative filtering algorithm into two types: one type is the collaborative filtering algorithm based on high similarity, and the other is the collaborative filtering algorithm based on the data model. The collaborative filtering method based on the high similarity of neighbors focused on analyzing the relationship between users and entities; the model-based collaborative filtering algorithm learned the prediction model through the user's rating of the entity or other behaviors, built the feature association between the user and the entity, and recommended the entity with high user preference to the users [2]. Many scientists had studied collaborative filtering algorithms based on data models, such as Ngaffo AN's research on matrix factorization factor models. The matrix factorization factor model is widely used. The matrix factor algorithm is used to extract hidden factors from the evaluation matrix of user objects, and these factors are used to describe users and

entities, as well as users' estimated values of other entities [3]. Besides, Angadi et al. discussed Naive Bayes model. The model had a solid mathematical theory and steady-state classification efficiency as the basis, and the algorithm required by the model was relatively simple, and the main content of the algorithm was to solve some parameters with incomplete data [4]. In addition, there is an analysis of the hidden semantic analysis model performed by Cai. The model used statistical econometrics to analyze large amounts of text, extracted implicit semantic structures that exist between words, and used this kind of implicit semantic structure to reconstruct sentences [5].

Problems of the current recommendation algorithm: (1) the problem of cold start. A study by Villanueva-Polanco and Angulo-Madrid pointed out that a cold start is when a new entity that has never been evaluated is added to the system and its associated data is extremely scarce. In addition, when a newly registered user joins the system, the user's data is extremely lacking, and it is also impossible to recommend [6]. Karacan et al. proposed and applied a novel overlapping method that employs overlapping techniques as a tool to deal with the shortcomings of clustering techniques. The advantage of overlapping techniques is to allow users to feedback their behavior and ratings in social networks, belonging to multiple clusters simultaneously [7]. (2) The problem of sparsity: Chen et al. combines the evaluation of individual cognitive behaviors, user cognitive relationships, and time decay coefficients into a probability matrix decomposed by a single model, combined with the social interaction coefficient for personalized recommendation, for the use of social interaction factor for personalized recommendations [8]. Rodpysh et al. pointed out that the reason for the sparse data is that most users have a very limited number of reviews for the product. Second, as the total number of products and users increases, the number of products viewed by users will only decrease. The proportion of evaluation items to the total is even lower [9]. The sparsity problem can cause inefficiency, low precision, and low adequacy for similarity computation. (3) Real-time questions: according to the research of Paddock, the total number of items and users for calculating similarity keep increasing, so when all the sparse matrices are passed, the time spent increases exponentially, and users' demands for real-time recommendation will be difficult to meet [10].

At present, the research on the recommendation method of tourist attraction recommendation platform had achieved certain results, but there still had a problem about poor recommendation accuracy. This paper applied big data and artificial intelligence algorithms to the improvement of recommendation algorithms, and payed attention to solve the problem of low recommendation accuracy. This paper can accurately discover the needs of users, help users to obtain recommendation results efficiently and accurately, and meet the personalized needs of users.

## 2. Tourist Attraction Recommendation System

The tourist attraction recommendation system can enable users to obtain attractions that match their preferences, save

a lot of time for users, and improve users' satisfaction with the platform, thereby increasing the visit rate of the tourist attraction recommendation platform and enhancing its market competitiveness. The recommendation system is based on the recommendation algorithm. Firstly, the user data is extracted from the big data, and the data is simply analyzed and interpreted. Then, the data is mined, and the user model is generated. Finally, the hybrid recommendation algorithm improved by the artificial intelligence algorithm module completes the prediction of the user's scenic spot preference. It is shown in Figure 1.

*2.1. Big Data Module.* Because the data collection of big data is very large, the requirements for data retrieval, storage, control, and analysis are very high [11]. It has four characteristics of large scale, fast flow, diverse types, and low value density. The commonly used big data technologies are actually based on the Hadoop ecosystem [12]. Hadoop uses a decentralized system infrastructure. In other words, its data storage and processing processes are distributed and completed by multiple machines. Through such parallel processing, security and data processing scale are improved. The main designs of the Hadoop platform: HDFS and MapReduce [13]. HDFS provides massive storage, and MapReduce provides massive computing. When submitting computing tasks to the MapReduce platform, the computing tasks are first divided into different partitions, each partition processes some input data and then sinks to different nodes. When the Map task is completed, Reduce will combine the outputs of several previous maps and then output. It is equivalent to using distributed machines to complete large-scale computing tasks.

Therefore, this paper drew inspiration from the Hadoop ecosystem and built a big data module for building a tourist attraction recommendation system. In this system, the data sources of users and attractions are extensive and diverse, and a complex data environment is formed in parallel, which brings a problem that must be solved for big data processing. Therefore, the big data module must obtain, clean, and integrate the required data sources in order to obtain the essence and relationship of the data. After a series of reconstructions, the data is emptied and stored in a similar structure to ensure high data quality and reliability. The module composition is shown in Figure 2.

Data extraction is not a brand new technology and has been relatively mature in the traditional database field [14]. Data retrieval is the process of filtering the desired data using specific criteria, transferring the data to a target file, and searching the entire data source. As the first step of data processing, data mining plays an important role.

The principle of data cleaning is to use the existing technical means to transform the low-quality data in the source data into data that can meet the quality requirements by analyzing the existing format and the reasons for redundant data, so as to ensure the data quality of the datasets generated by the later data integration [15]. Data integration integrates scattered but interrelated data into a unified dataset according to a certain logic, so that the unity and utilization efficiency of the overall data can be guaranteed [16]. Data

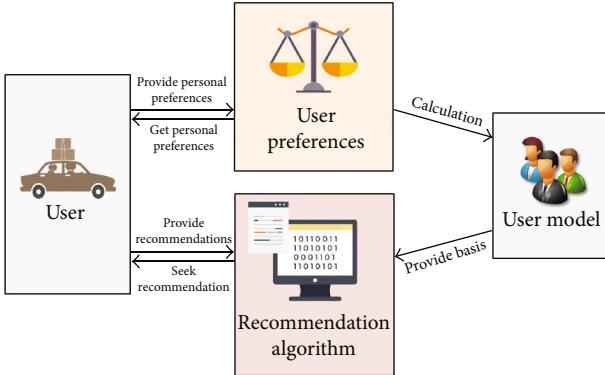


FIGURE 1: Tourist attraction recommendation system.

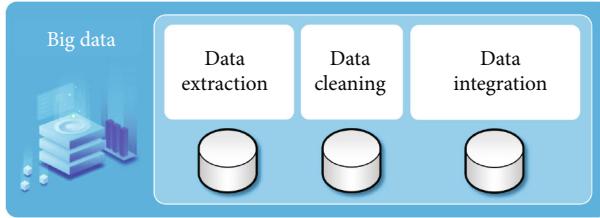


FIGURE 2: Big data module.

integration can provide comprehensive data sharing and provide an integrated data source access interface that is convenient for other modules to access the big data module.

**2.2. Engine Service Module.** Engine service module consists of data mining group and user model. The reason it is called “engine service” is that this module is at the heart of the whole system [17]. Its final result will produce a related user model, which contains various data of the user: age; gender; travel time; travel footprint; travel mode of transportation; distance between tourist attractions and residence. The data will be compiled into three types of attributes: interest preference, explicit feedback, and implicit feedback. Only when the generated user model data is accurate can the artificial intelligence algorithm make accurate recommendations; otherwise, it is water without a source and a tree without roots. The module composition is shown in Figure 3.

In this module, the data mining group will reanalyze the data processed by the big data module and generate a user model that includes three attributes: interest preference, explicit feedback, and implicit feedback.

**2.3. Artificial Intelligence Algorithm Module.** Artificial intelligence has become a broader term nowadays. In the past, a large number of the term artificial intelligence is regularly used interchangeably with subfields such as machine learning and deep learning. However, they differ from each other in a number of ways. Machine learning, for example, focuses on building systems that can learn or improve performance based on the data they use [18]. In other words, all machine learning is artificial intelligence, but not all artificial intelligence is machine learning. In order to give full play to the value of artificial intelligence, scholars have gradually deepened their research on data science. Data science is an inter-

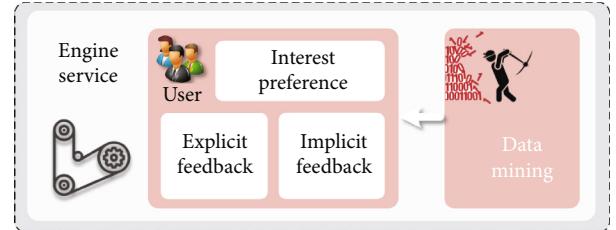


FIGURE 3: Engine service module.

disciplinary field that combines professional skills and knowledge from science, data statistics, computer science, and other disciplines, supplemented by other methods to extract data value and conduct a comprehensive and detailed study of data collected and generated from multiple sources analysis [19].

The artificial intelligence algorithm module studied in this paper includes four aspects: data processing, feature engineering, model evaluation, and machine learning. The focus of this module is machine learning, mainly generating data models through Bayesian networks and improving hybrid recommendation algorithms through artificial intelligence algorithms. The module composition is shown in Figure 4.

### 3. Algorithm for Tourist Attraction Recommendation

**3.1. Introduction to Bayesian Networks.** Bayesian network is a graphical model whose operations contain probabilistic content, so it is often used to analyze and predict the relationship between two objects [20]. As shown in Figure 5, it is a classic Bayesian network diagram. If the probability of  $X_4$  is calculated, it needs to rely on  $X_1$ ,  $X_2$ , and  $X_3$ . It can be seen that the Bayesian network involves probability problems. The following is a preliminary introduction to Bayesian probability.

The Bayesian formula is a conditional probability formula, such as  $P(A|B)$ , which represents the probability of  $A$  under the condition of  $B$ , where  $P(A|B)$  represents the probability of  $A$  and  $B$  occurring at the same time. From this, the calculation formulas of  $P(A|B)$  and  $P(B|A)$  can be obtained as follows:

$$P(A|B) = \frac{P(A, B)}{P(B)}, \quad (1)$$

$$P(B|A) = \frac{P(A, B)}{P(A)}. \quad (2)$$

The formula for calculating the Bayesian posterior probability after processing is as follows:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}. \quad (3)$$

**3.2. Bayesian Algorithm.** Bayes’ theorem is often used to calculate the probability of a condition occurring by

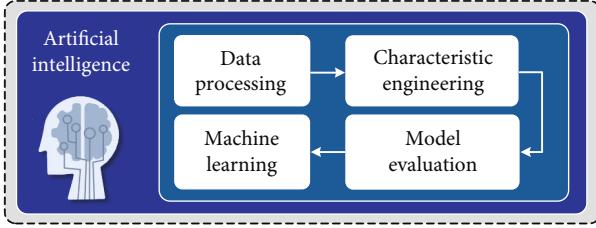


FIGURE 4: Artificial intelligence algorithm module.

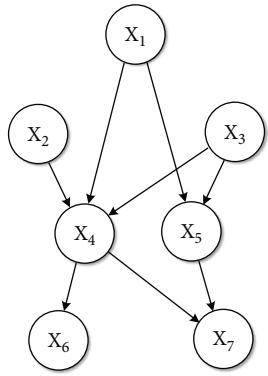


FIGURE 5: Bayesian network.

quantifying the relationship between two objects. For example, let  $X$  be the head precursor data, which contains unknown attributes whose value is  $(x_1, x_2, \dots, x_n)$ . For the classification problem, after  $X$  is determined, the probability of assuming that  $H$  is established is  $P(H|X)$ , which means that under the condition  $X$  is the posterior probability of  $H$ . Bayes' Theorem provides a method consisting of  $P(H)$ ,  $P(X)$ , and  $P(X|H)$ , similar to Formula (3), where  $P(H|X)$  operates as follows:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}. \quad (4)$$

$P(H)$  is the prior probability of  $H$ ,  $P(X|H)$  is the posterior probability of  $X$  under condition  $H$ , and  $P(X)$  is the prior probability of  $X$ . Bayes satisfies the condition that each feature data in a special set is independent of each other. And from the probability theory, it can be concluded that when  $A$  and  $B$  are independent of each other,  $P(AB) = P(A)P(B)$ . The calculation formula of its  $P(X|H)$  is as follows:

$$P(X|H) = \prod_{k=1}^n P(x_k|H) = P(x_1|H)P(x_2|H) \cdots P(x_n|H). \quad (5)$$

**3.3. Application of Bayesian Network in Scenic Spot Recommendation.** In the stage of predicting user preference, it is necessary to calculate the user's preference for an entity. The probability formula can be derived to express the user's preference for an entity, and a data model can be evaluated

and generated. The generation process is as follows: first, define the user set and the scenic spot set: define the user model as a set, which is used to represent  $n$  users, and the set is  $U = \{u_1, u_2, \dots, u_n\}$ . Among them,  $u_i (i \in n)$  represents the user feature vector, and all the features are aggregated to represent a user feature vector. The scenic spot set is defined as  $A = \{a_1, a_2, \dots, a_m\}$ , which represents  $m$  users, where  $a_{ii \in m}$  represents the scenic spot feature vector. All features collectively represent an attraction record. The best choice is obtained by improving formula (5). The calculation formula of the user's preference for the scenic spot is as follows:

$$P(a_i|u_j) = \frac{P(a_i)P(u_j|a_i)}{P(u_j)}. \quad (6)$$

**3.4. User Model Generated by Bayesian Network.** The formula for calculating the user's preference for scenic spots generated by the Bayesian network is reanalyzed to generate a user model.

Definition  $P(a_i|u_j)$  represents the probability of recommending a certain scenic spot to a user, that is, the user's scenic spot recommendation degree. The attribute coincidence between users is defined, which is used to judge the similarity of users' preference attributes for scenic spots. The calculation formula is as follows:

$$\text{Sim}_b(u, k) = \frac{N_{fu} \cap N_{fk}}{N_{fu} \cup N_{fk}}. \quad (7)$$

Define  $\text{Sim}_b(u, k)$  to represent the similarity of behavior between users, and its calculation formula is as follows:

$$\text{Sim}_b(u, k) = \frac{\sum_{a \in c_{u,k}} Ns_{u_a} Ns_{k_a}}{\sqrt{\sum_{a \in c_u} Ns_{u_a}^2} \sqrt{\sum_{a \in c_k} Ns_{k_a}^2}}. \quad (8)$$

$C_{u,k}$  represents the attractions visited by both user  $u$  and user  $k$ ,  $N_{s_{u_a}}$  represents the number of check-ins by user  $u$  at attraction  $a$ ,  $N_{s_{k_a}}$  represents the number of times user  $k$  has checked in at attraction  $a$ ,  $C_u$  is the number of attractions that user  $u$  has checked in, and  $C_k$  is the number of attractions that user  $k$  has checked in number. Definition  $\text{Sim}(u, k)$  represents the similarity between user  $u$  and user  $k$ , including user behavior similarity and user attribute similarity. The calculation formula is as follows:

$$\text{Sim}(u, k) = b\text{Sim}_b(u, k) + (1 - b)\text{Sim}_f(u, k). \quad (9)$$

$b$  is a weighting factor used to adjust the similarity of user behavior and the similarity of user attributes, and the proportion of a certain factor is increased according to the actual situation. If the user is a new user without any information record, then define  $b = 0$  at this time, thus solving the problem that the new user has no data under

the cold start problem. Define  $R_{u,a}$  to represent user u's rating for attraction a, and its calculation formula is as follows:

$$R_{u,a} = \gamma \frac{N_{u,a}}{\sum_{j \in A} N_{u,j}}. \quad (10)$$

$N_{u,a}$  represents the number of check-ins of user u at attraction a,  $\sum_{j \in A} N_{u,j}$  represents the number of check-ins of user u, and A table indicates the set of user attractions, and  $\gamma$  is the adjustment factor to prevent the value of  $R_{u,a}$  from being too small. So far, the user model has been created.

**3.5. Introduction to Leapfrog Algorithm.** Hybrid leapfrog algorithm be called SFLA for short belongs to artificial intelligence algorithm. The "frog" group is initialized, and information is transmitted according to the differentiated groups, and the global optimal solution of the problem is obtained by combining the deep subgroup search function and the overall global information exchange function of the frog group. The SFLA algorithm not only has the faster global optimization capability of the particle swarm optimization algorithm but also has the local optimization capability and can perform the evolutionary algorithm faster. The algorithm has the advantages of convenient implementation, fast optimization, and strong global search ability and has became one of the most popular research fields for researchers. The concept diagram of hybrid leapfrog search is shown in Figure 6.

In the SFLA algorithm, group of frogs is composed of several frog individuals, and each group is divided into several subpopulations containing a random number of frogs. The data of frog individuals in different subpopulations are different, and each frog in the subpopulation has certain special data, which will affect other individuals. Then, after the local search is performed, data is exchanged in each subpopulation, and the local optimal solution can be found in the subpopulation.

**3.6. Improved Hybrid Recommendation Algorithm Based on Artificial Intelligence.** The SFLA algorithm in the artificial intelligence algorithm and the collaborative filtering algorithm are introduced for fusion and improvement to generate a hybrid recommendation algorithm. The mathematical model of the algorithm can be interpreted as: all kinds of substitutes and measurement relations used by the algorithm during the execution process will automatically organize the data into a mathematical model, which is conducive to the analysis of the algorithm. Next, analyzing and explaining the mathematical model of the SFLA algorithm: randomly generate  $P$  frog individuals to form the initial frog population  $Q = \{F_1, F_2, \dots, F_p\}$ , and the  $i$ th frog in  $Q$  is denoted as  $F_i \{x_{i1}, x_{i2}, \dots, x_{il}\}$ . This formula represents a solution to the problem, where  $l$  refers to the dimension of the solution space. Arranging all the frogs in the population in descending order of fitness and dividing the population into  $m$  subpopulations, each subpopulation has  $n$  frogs, and the initial

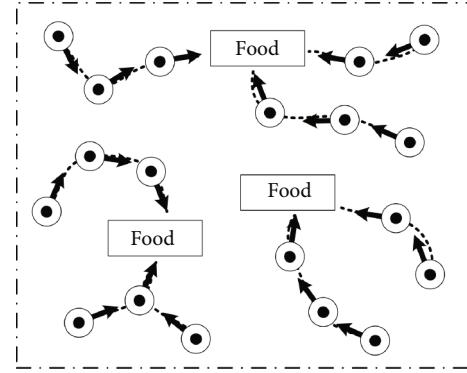


FIGURE 6: Hybrid leapfrog search.

total number  $P$  of frogs satisfies  $P = m \times n$ . When the population is divided into subpopulations, the rules described by the formula are used to divide, where  $F^k$  refers to the  $k$ th frog subpopulation.

$$F^k = \left\{ F_{k+m(j-1)} \in P \mid 1 \leq j \leq n \right\}, \quad 1 \leq k \leq m. \quad (11)$$

Local search is carried out in each subpopulation and iteratively updated, that is, the position of the frog  $F_w$  with the worst fitness in each subpopulation is continuously updated iteratively. The updated rules are as follows:

$$D_i = \text{rand}(0, 1) \times (F_b - F_w), \quad (12)$$

$$\text{new}F_w = F_w + D_i - D_{\max} \leq D_i \leq D_{\max}. \quad (13)$$

In the formula,  $\text{rand}(0, 1)$  refers to a random number between 0 and 1, while  $D_i$  refers to the distance movement value in the  $i$  direction, and  $D_{\max}$  refers to the upper limit of the single position movement of each frog. After the location update operation is performed, if the fitness of  $\text{new}F_w$  is better than that of  $F_w$ , use  $\text{new}F_w$  instead of  $F_w$ . The fitness of  $\text{new}F_w$  is not as good as that of  $F_w$ , replacing  $F_g$  with  $F_b$  and performing the update operation again.

The normal distribution theory is introduced to solve the optimization problem of local minima. The basic theory of normal distribution is as follows: a random variable  $x$  that obeys the normal distribution of mathematical expectation is  $u$ , and the variance is  $\sigma^2$  which can be expressed as  $N(u, \sigma^2)$ , and the standard normal distribution is the normal distribution when  $\sigma = 1$ ,  $u = 0$ . Its probability density function is as follows:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-u)^2}{2\sigma^2}\right). \quad (14)$$

A variation factor that follows a normal distribution is added to the update strategy of the worst frog, where  $N(0, 1)$  refers to the standard normal distribution. The improved strategy is as follows:

$$\text{new}F_w = F_w + N(0, 1) \times D_i - D_{\max} \leq D_i \leq D_{\max}. \quad (15)$$

The mutation strategy of the fittest is added to the SFLA algorithm to improve the lower limit of the quality of the population. The addition of the variation factor can reduce the blindness of optimization, thereby improving the execution speed of the algorithm.

$$\text{new}F_{mij} = F_{mij} + N(0, 1)_{mij} \times F_{mij}. \quad (16)$$

On this basis, the winning mutation mechanism is added to make the population move towards a better solution, and it is easier to find the optimal solution. The winning mutation strategy can be expressed as:

$$\text{new}F_{bj} = F_{bj} * \left(1 + \xi N(0, 1)_{bj}\right). \quad (17)$$

The similarity value of the target user  $u$  and the  $U'$  element in the nearest neighbor set can be used as the fitness function, and the final expression of the algorithm is obtained as follows:

$$f(F_k) = \sum_{n=1}^N x_{kn} \text{sim}(u, u'). \quad (18)$$

In order to verify whether the algorithm is effective, two high-dimensional unimodal functions, the Sphere function and the Rosenbrock function, are selected to test the performance of the algorithm. The detailed mathematical formulas and characteristics of the functions are as follows.

The Sphere function is often used to measure the accuracy of an algorithm. In the value range, the function only obtains the globally unique optimal value 0 when the point is  $(0, 0, \dots, 0)$ . That is to say, the optimal value of the function is  $f(0, 0, \dots, 0) = 0$ .

$$f_1(X) = \sum_{i=1}^n x_i^2, x_i \in [-100, 100]. \quad (19)$$

The Rosenbrock function has the characteristics of non-convex and asymmetric and is generally used to measure the operating efficiency of the algorithm. The function only obtains the globally unique optimal value 0 when the point is  $(1, 1, \dots, 1)$ . In other words, the optimal value of the function is  $f(1, 1, \dots, 1) = 0$ .

$$f_2(X) = \sum_{i=1}^{n-1} \left[ (x_{i+1} - x_i^2)^2 + (x_i + 1)^2 \right], x_i \in [-30, 30]. \quad (20)$$

#### 4. Data Sources

Taking the data obtained online as a dataset, records of 1000 users and 2851 tourist attractions were obtained. User attributes include user ID, gender, age, location, and check-in time; scenic spots include scenic spot identification, user ID, scenic spot description, scenic spot location, scenic spot label, and check-in times. The process of processing the dataset is as follows: first, the operation of prescreening

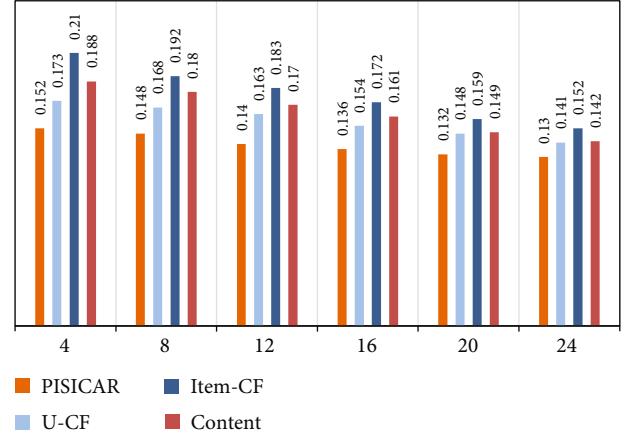


FIGURE 7: The effect of the number of adjacent users on the mean absolute error.

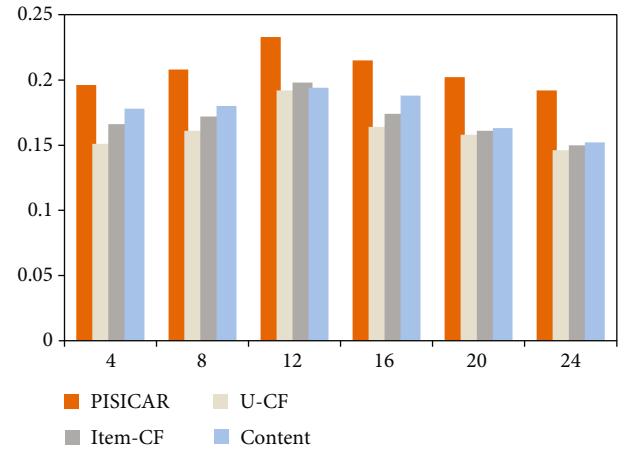


FIGURE 8: Accuracy comparison.

and processing the dataset is carried out; the purpose is to decompose the users who generate a very low number of travel records, so as to improve the accuracy of recommendation. Then, according to the ratio of 8 to 2, the dataset is divided into training set and test set. The training set is used to construct the tourist attraction recommendation algorithm model, and the test set is used to verify the accuracy of the tourist attraction recommendation algorithm.

#### 5. Experiment of Recommendation Algorithm

The improved algorithm was named PISICAR, performing experimental analysis on the dataset after processing the data required for the experiment. According to the proposed tourist attraction recommendation algorithm and experimental analysis, it is concluded that the mean absolute error of the algorithm recommendation results is associated with the amount of relevant factors selected as users in the experiment. Therefore, this experiment accepted a different number of adjacent user selections and used PISICAR as a collaborative user filtering-based attractions recommendation algorithm, as well as element-based collaborative filtering and content-based algorithms used in the comparative

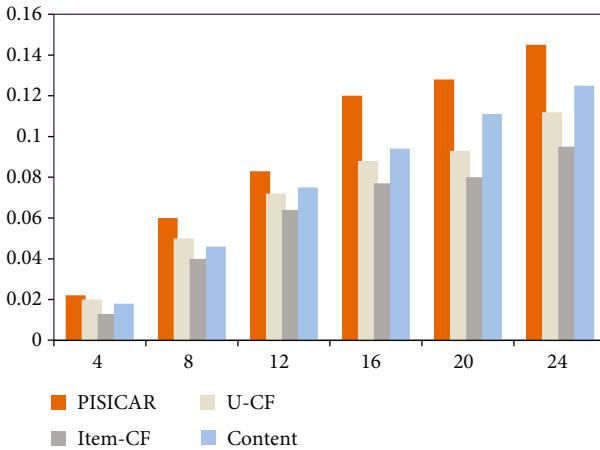


FIGURE 9: Recall comparison.

experiments for comparison. U-CF is used to present a recommendation algorithm for tourist attractions based on joint user filtering, Item-CF is used to present a recommendation algorithm for filtering objects of joint tourist attractions, and Content is a content-based tourism used to present the details of attractions. This paper presents the strength of the algorithm performance by calling the parameters of the mean absolute error value. The specific experimental results are shown in the figure.

Neighboring user sets are called for comparison of mean absolute error values in the experiments. The number of its calls is different, and the corresponding mean absolute error value will also change, but the whole is gradually stable with the amount of the growth of the users of the mean absolute error value. By comparing the difference of the number of adjacent user parameters of different algorithms in Figure 7, the average absolute error value of the recommendation algorithm proposed in this paper is much lower than that of the other three algorithms.

The accuracy comparison is shown in Figure 8.

The recall comparison is shown in Figure 9.

From Figures 8 and 9, the precision rate and the recall rate increase with the number of recommendations. However, the accuracy rate reached an inflection point when the number of recommendations increased to a certain extent and then showed a downward trend. It is shown that the accuracy and recall of the algorithm change when the number of tourist attractions recommended to the user changes. A comparison of the algorithm proposed so far in this paper with three other commonly used unimproved algorithms shows that the algorithm proposed in this paper has higher precision and recall. Further analysis showed that the recommendation accuracy of this algorithm is 6.55% higher than that of the unimproved algorithm.

## 6. Conclusion

For a travel recommendation platform, a higher recommendation accuracy rate can bring users a better user experience. Based on big data and artificial intelligence algorithms, this paper improved the tourist attraction recommendation algo-

rithm. There were still shortcomings in the research on the recommendation algorithm of tourist attractions. There are several aspects that can be improved:

- (1) In the aspect of recommendation algorithm research, the method used to formulate label parameters for the user model lacks pertinence, and the multifaceted behavior patterns of users cannot be added to the user model. Also, there are not enough labels to cluster sights. Subsequent research will optimize these two aspects
- (2) In this paper, the research depth of tourist attraction method based on big data and artificial intelligence algorithm is not enough. The next step is aimed at addressing new users and new attractions. In addition, it is also necessary to introduce scenario factors for scenario analysis to provide extensive and in-depth support for the recommendation of tourist attractions
- (3) The big data processing platform still needs to be improved, and there is still a lot of room for improvement in the processing mechanism in the face of diverse and complex data. The next step is to optimize the data algorithm and add it to the data warehouse to better carry out the work of online analysis

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

## Acknowledgments

This work was supported by the Scientific Research Project of Colleges and Universities in Hainan Province (grant no. Hnky2021ZD-26, Research on key technologies of student credit investigation and certificate deposit based on blockchain), Hainan Provincial Natural Science Foundation of China (grant no. 621RC1082), and Research Fund Project of Qiongtai Normal University (qtyb201810).

## References

- [1] Y. Liu, M. Yin, and X. Zhou, "A collaborative filtering algorithm with intragroup divergence for POI group recommendation," *Applied Sciences*, vol. 11, no. 12, pp. 5416–5416, 2021.
- [2] X. Han, Z. Wang, and H. J. Xu, "Time-weighted collaborative filtering algorithm based on improved mini batch K-means clustering," *Advances in Science and Technology*, vol. 105, no. 2, pp. 309–317, 2021.
- [3] A. N. Ngaffo, W. E. Ayeb, and Z. Choukair, "Service recommendation driven by a matrix factorization model and time series forecasting," *Applied Intelligence*, vol. 52, no. 1, article 2478, pp. 1110–1125, 2022.

- [4] U. B. Angadi, A. Rai, and G. Uma, "MBFerns: classification and extraction of actionable knowledge using multi-branch ferns-based Naive Bayesian classifier," *Soft Computing*, vol. 25, no. 13, pp. 8357–8369, 2021.
- [5] A. Cai, "A latent semantic analysis-based image tag optimisation method," *International journal of applied decision sciences*, vol. 13, no. 1, pp. 109–121, 2020.
- [6] R. Villanueva-Polanco and E. Angulo-Madrid, "Cold boot attacks on the supersingular isogeny key encapsulation (SIKE) mechanism," *Applied Sciences*, vol. 11, no. 1, article app11010193, pp. 193-194, 2021.
- [7] H. Karacan, H. Karacan, and Y. E. Yenice, "A novel overlapping method to alleviate the cold-start problem in recommendation systems," *International Journal of Software Engineering and Knowledge Engineering*, vol. 31, no. 9, pp. 1277–1297, 2021.
- [8] R. Chen, Y. S. Chang, Q. Hua, Q. Gao, X. Ji, and B. Wang, "An enhanced social matrix factorization model for recommendation based on social networks using social interaction factors," *Multimedia Tools and Applications*, vol. 79, no. 19-20, pp. 14147–14177, 2020.
- [9] K. V. Rodpysh, S. J. Mirabedini, and T. Banirostam, "Correction to: employing singular value decomposition and similarity criteria for alleviating cold start and sparse data in context-aware recommender systems," *Electronic Commerce Research*, vol. 22, no. 1, article 9497, pp. 223–223, 2022.
- [10] A. Paddock, "Real-time data," *International Journal of Computer Applications*, vol. 175, no. 32, pp. 28–31, 2020.
- [11] L. Cui, "Construction of big data technology training environment for vocational education based on edge computing technology," *Wireless Communications and Mobile Computing*, vol. 2022, 9 pages, 2022.
- [12] E. V. Nikulchev, A. V. Tatarintsev, and V. Belov, "Choosing a data storage format in the Apache Hadoop system based on experimental evaluation using apache spark," *Symmetry*, vol. 13, no. 2, pp. 195–198, 2021.
- [13] A. Smm, A. As, and B. Ant, "Cross-map reduce: data transfer reduction in geo-distributed map reduce," *Future Generation Computer Systems*, vol. 115, no. 3, pp. 188–200, 2021.
- [14] N. E. Zamri, M. A. Mansor, and M. Kasihmuddin, "Amazon employees resources access data extraction via clonal selection algorithm and logic mining approach," *Entropy*, vol. 22, no. 6, pp. 596-597, 2020.
- [15] A. M. Venkatachalam, A. Perera, S. E. Stutzman, D. W. M. Olson, V. Aiyagari, and F. D. Atem, "Methods for cleaning and managing a nurse-led registry," *Journal of Neuroscience Nursing*, vol. 52, no. 6, pp. 328–332, 2020.
- [16] J. Zhu and P. Wu, "Towards effective BIM/GIS data integration for smart city by integrating computer graphics technique," *Remote Sensing*, vol. 13, no. 10, pp. 1889–1895, 2021.
- [17] J. Ong, A. Selvam, and J. Chhablani, "Artificial intelligence in ophthalmology: optimization of machine learning for ophthalmic care and research," *Clinical and Experimental Ophthalmology*, vol. 49, no. 5, pp. 413–415, 2021.
- [18] K. S. Lee, I. S. Song, E. S. Kim, H. I. Kim, and K. H. Ahn, "Association of preterm birth with medications: machine learning analysis using national health insurance data," *Archives of Gynecology and Obstetrics*, vol. 305, no. 5, pp. 1369–1376, 2022.
- [19] J. Torres, A. Zalapa-Damián, and F. J. Domínguez-Mota, "Data science and machine learning technique for predicting electrical resistivity in recycled concrete with nopal as addition," *Forum*, vol. 40, no. 8, pp. 43–62, 2021.
- [20] Z. Duan, L. Wang, and M. Sun, "Efficient heuristics for learning Bayesian network from labeled and unlabeled data," *Intelligent Data Analysis*, vol. 24, no. 2, pp. 385–408, 2020.