ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Optimization of Precision Marketing Strategies under Big Data Analysis of User Behavior in Smart Tourism Platforms

Aifang Zhang[1,*], *Lingling Zhang*[2] *and Zhengzheng Zhang*[3]

[1] School of Tourism Management and Rural Revitalization, Luoyang Vocational College of Culture and Tourism, Luoyang, Henan, 471000, China
[2] The Institute of Marxism, The School of Philosophy, Law & Politics, Shanghai Normal University, Shanghai, 200234, China
[3] Gongyi Xicun Second Primary School Dongcun Village, Gongyi, Henan, 451281, China

SUMMARY: *K-means clustering algorithm is employed in this paper to group preprocessed user behavior data and identify appropriate clustering centers. To enhance the performance and precision of the clustering findings, a clustering algorithm that uses chaotic ant colony optimization approach (CAS-C) is proposed. The association rule algorithm is applied to discover the association rules among various clusters to study the correlation between user behavior and attraction interests. Subsequently, the pattern recommendation algorithm is applied to compute the association weight of every attraction and suggest available attractions to target users. Six classes are used to cluster the users and the lowest support and the lowest confidence of mined association rules are greater than 65%. The calculation results of core attractions and routes with user interest show that the best recommendation effect is obtained at the point where the number of single recommendations and is 30.*

KEYWORDS: *k-means clustering; CAS-C; association rule mining; pattern recommendation; smart tourism*

1 Introduction

In today's increasingly competitive global tourism market, big data has become a core asset for enterprises to understand the customer base, grasp the demand, and improve the marketing effect [1]. The application of smart tourism platform has a high degree of data diversity: tourists in different regions, travel needs in different time periods, information acquisition habits in different channels, and linked purchasing behaviors of various types of accommodations, transportation, and attractions [2, 3]. By transforming these seemingly messy data, behaviors, and preferences into clear insights, and then into precise marketing actions, it is a key path to achieve efficient placement, increase unit price, and enhance user stickiness [4, 5].

Smart tourism platforms can collect various behavioral data of tourists, including browsing records, consumption habits, travel preferences, etc., so as to accurately understand the needs of each tourist [6]. Tourism enterprises can provide tourists with personalized services based on these data, such as recommending local museums and monuments for tourists who like history and culture, arranging special restaurant tours for food-loving tourists, and designing exclusive play routes and activities for parent-child families [7-9]. In this way, tourists can get a tourism experience more in line with their own mind, and satisfaction will be greatly improved

*zhang_198134@163.com
https://doi.org/10.65102/is2026447 Published: 30 April, 2026

[10]. And in the fierce market competition, who can better realize the precise marketing strategy will stand out. Under the big data analysis of user behavior in the smart tourism platform, the tourism service process is optimized through the use of new technologies to improve operational efficiency and reduce costs [11, 12]. For example, the online booking system can reduce manual operation and improve the booking processing speed; intelligent tour guide equipment can save tour guide manpower, while providing tourists with richer explanation content. Moreover, with the big data analysis of user behavior on the smart tourism platform, enterprises can also understand the market dynamics and competitors, adjust marketing strategies, and launch more attractive products and services, so as to occupy a favorable position in the competition [13-15].

Precision marketing based on the analysis of user behavior in the context of smart tourism will help to reduce the costs and increase the efficiency, enhance the experience of tourists and eventually achieve a win-win scenario. Literature [16] examines the state of research on the analysis of user behavior and the precision marketing strategy in smart tourism, and also studies the implementation of the precision marketing strategy in smart tourism with the aim of providing the data-driven decision-making support to tourism companies using data. Literature [17] will focus on examining and optimizing personalized travel service recommendation algorithm to enhance the accuracy of recommendations made and user satisfaction of travel services and by introducing a collaborative filtering recommendation algorithm based on user behavior analysis, the algorithm is significantly improved in terms of recommendation accuracy and user experience. Literature [18] introduced a user behavior trajectory framework in the context of precision marketing in tourism and built a tourism precision recommendation system that can target tourism groups and make tourism smarter. Literature [19] highlights the importance of personalized marketing and user engagement content in electronic word-of-mouth marketing and recommends that tourism companies should encourage their customers and employees to publish some tourism-related content on electronic media. Literature [20] highlighted the significance of the analysis of tourism consumer behavior by big data in the precision marketing of scenic spots and explored the association between different segments of big data and the amount of tourists in scenic spots, along with how the consumption behavior of secondary consumption items in scenic spots differs and what effects they have, in order to find out the potential of the growth of the scenic spot businesses. Literature [21] describes the use of artificial intelligence in the tourism sector to enable the creation of digital marketing strategies in the tourism sector, a novel business ecosystem, which is capable of analyzing and extracting large volumes of user behavioral data to use in its precise marketing strategies. Literature [22] addresses the transformational change brought about by big data analytics in the tourism sector especially the possibilities created by this technology in regards to improving the customer experience, streamlining operations, and making strategic decisions, and the issues that need to be considered to fully utilize such advancements. Literature [23] describes the use of artificial intelligence technologies in the hospitality industry to effectively analyze customer behavior using customer behavior and hence inform the use of optimized customer loyalty strategies.

In this paper, users' behavioral data such as hotel booking and travel information in the smart tourism platform are collected and used as in-depth research objects after completing invalid data cleaning and type conversion. Through K-means clustering algorithm to calculate the average value of each behavioral data distance from the class center, to complete the initial clustering. In order to improve the clustering and subsequent marketing effect, CAS-C algorithm is introduced to solve the clustering error sum of squares (SSE) objective function, and the global iteration searches for the best class center point to determine the number of clusters. Combined with the association rule algorithm to mine the hidden rules in the clustered data to ensure that the rules satisfy the minimum trust and support thresholds. Apply the recommendation algorithm based on user group visit patterns to extract the attractions in the valid rules and calculate their association weights with different categories of users to realize the marketing recommendation of attractions.

2 Behavioral data analysis based attraction recommendation implementation

2.1 Clustering Basis and Data Preparation

2.1.1 Cluster analysis

The fundamental algorithm used in clustering is the K-means clustering algorithm. It is a partition clustering algorithm using average value as the center of the class, it is simple, its clusters are straightforward to interpret, and it has the highest efficiency when dealing with large datasets, to which this paper can apply to address the issue of analyzing the behavior of the smart tourism platform users. Assume that there are N objects and they are split into K classes. Figure 1 represents the concept of K-means clustering.

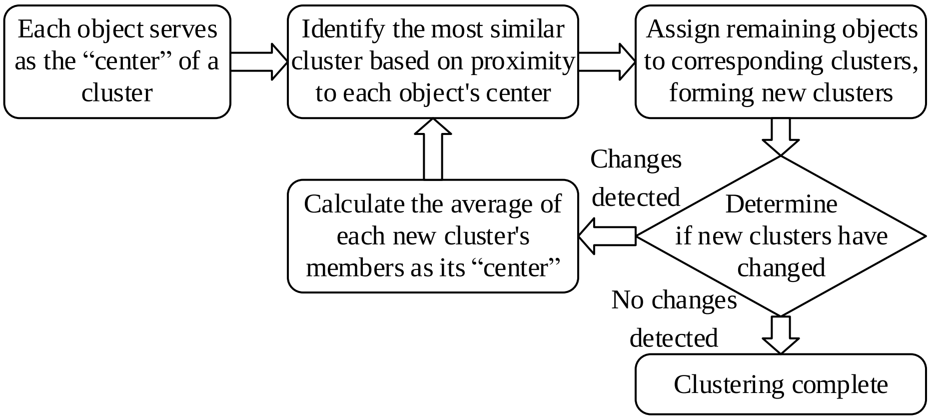

Figure 1: The Principle of K-means Clustering

Combined with the results of previous research, the clustering process in this paper is divided into the following six steps.

1) Definition of the problem: user behavior segmentation through the user behavior of smart tourism platform to find out the target users.

2) Preparation of data: collect user behavior data of a platform, go through the analysis of data, data pre-processing, and data standardization, so as to meet the data requirements of clustering.

3) Selection of clustering variables: Comparative study and analysis of domestic and international indicators of tourism user segmentation on smart tourism platforms, combined with the variables of clustering data, to determine the clustering variables and descriptive variables.

4) The fourth one is the determination of the number of clusters: The approach offered by SPSS Clementine can be used to determine the number of clusters based on the real characteristics of the user of the behavioral data.

5) Clustering of users using SPSS.

6) Analysis of clustering results, analyzing, evaluating, and interpreting the classification

of clusters.

2.1.2 Data collection and pre-processing

Cluster analysis is one of the methods of data analysis used in exploration. Typically, we apply cluster analysis to sort and classify apparently disorganized items so that we can gain more insight into the research item. The outcomes of clustering should have a high similarity between objects in the same group and a low similarity among objects in different groups.

1) Data collection. This study uses data from a smart tourism platform, and the survey object is all users who have used the platform program. In addition to the user's personal information username also designed to extract information on several attribute dimensions, in order of user level, total orders, room nights, user source, booking method, purchase frequency, help people booking. A total of 75000 sets of data were obtained.

2) Data pre-processing. To improve the efficiency of data analysis and the rationality of the user clustering results, it is essential to ensure that the data to cluster is accurate, concise and simple to manipulate, but much effort must be done to prepare the data prior to developing the user clustering model. Firstly, data purification. Data purification eliminates erroneous, null, incomplete and other data in the data source that do not satisfy the demands of user clustering. The level of objective data in the data source of the smart tourism platform is generally satisfactory, except that the user self-registered information, such as gender and age, has more issues, and these data are not involved in clustering, so they do not play a significant role in the outcomes. The next step is the handling of outliers. Outliers are values that are very distant in terms of values compared to the overall worth of data, which are frequently deceptive to the user clustering procedure, yielding inaccurate findings, and in some cases, incorrect ones. The outliers may be addressed through substitution of the mean or median of the data and in this case direct exclusion of the outliers has been applied.

3) The cluster analysis algorithm needs every field as a numeric value whereas various properties of an object may have different units of measurement such that the process of clustering would give unreasonable clustering outcomes. It is therefore required to convert the data type. In this paper, SPSS-Clementine has been used as the clustering tool and it focuses on data mining and it converts all the attribute fields into numerical types and then it defines and instantiates the variable types.

2.2 Clustering optimization based on CAS-C algorithm

2.2.1 Optimization models for clustering problems

The optimization model of the clustering problem is described below:

The mathematical model of a general clustering problem is as follows: let a sample dataset $S=\{x_1,x_2,\cdots,x_n\}$, and the number of clustered categories k, and the main purpose of clustering is to find k categories $C_1,C_2,\cdots,C_k$, which satisfy the following equation:

$$\begin{cases} \bigcup_{i=1}^{k} C_i = S; \\ C_i \cap C_j = \varnothing, i, j = 1,2,\cdots,k; i \neq j; \\ C_i \neq \varnothing, i = 1,2,\cdots,k. \end{cases} \tag{1}$$

Many clustering algorithms use the sum of squares of errors (SSE) as the objective function to evaluate the effectiveness of clustering. Determine the error with every data point, that is, the

Euclidean distance to the closest center of mass, followed by computing the total of the squares of the errors; we prefer the one with the smallest sum of squares of the errors because it implies that the prototype of the clusters (the center of mass) is a better representation of the center of the class. The SSE form is defined below:

$$SSE = \sum_{i=1}^{K} \sum_{x_j \in C_i} distance\left(x_j, z_i\right) \tag{2}$$

where z_i denotes the center (center of mass) of the class C_i, and x_j denotes the data clustered into C_i. The center of a category is generally the center of all points in this category, which can be expressed as the average value, and the formula for z_i is as follows:

$$z_i = \frac{1}{\left|C_i\right|} \sum_{x_j \in C_i} x_j, i = 1, 2, \cdots, k \tag{3}$$

In fact, the clustering task can be viewed as a process of determining the centers of k classes, i.e., the process of finding the center z_i of each class C_i, $i = 1, 2, ..., k$.

Based on this view, the mathematical model of clustering can be described as the following process: given a data set $S = \{x_1, x_2, ..., x_n\}$, the clustering result can be determined as follows:

$$C_i = \left\{ x_j \left| \left\| x_j - z_i \right\| \le \left\| x_j - z_p \right\|, p \ne i, p = 1, 2, \cdots, k, x_j \in S \right. \right\} \tag{4}$$

where $\left\| \cdot \right\|$ denotes the distance between two data points, it can be seen from Eq. (4) that C_i consists of a set of data elements closest to the class center z_i.

The clustering problem can thus be viewed as a search for the k class centers $z_1, z_2, ..., z_k$. Define this purpose as an objective function as follows:

$$f = \arg \min_{z_1, z_2, ..., z_k} \sum_{j=1}^{n} \arg \min_{1 \le p \le k} \left\| x_j - z_p \right\| \tag{5}$$

where $\left\| x_j - z_p \right\|$ is the distance between a data point and its closest class center. Clustering process is to identify the k class centers which minimize this objective function.

2.2.2 Principles of the CAS-C Algorithm

Once we have mathematically modeled the clustering problem, we realize that the clustering problem can be seen as one of the classes of optimization problems, where the objective of optimization is to obtain a globally optimal solution of the objective function of the form (2). This paper offers a clustering algorithm which uses the chaotic ant colony optimization method (CAS-C) to address the above function optimization problem through analogy with the clustering process and the ant foraging process.The center of each class can be regarded as the food source that the ants are looking for. It should be noted that in CAS-C algorithm, there is no need to set the initial center of each class artificially, so this algorithm can effectively overcome the limitation of “initial value sensitivity”.In the initialization stage, the ant-based algorithm chooses some random sample points in the data set as the starting positions of the ants, and after a number of iterations, all the ants have finally converged to some points in the data space,which are regarded as representing the center of each class. The algorithm terminates when it is executed to the pre-set maximum number of iteration steps and divides the dataset according to the obtained clustering centers to give the final clustering results.

The center of each class to be found is denoted by $C_p\left(p=1,2,\cdots,k\right)$, k denotes the number of classes to be clustered, this parameter is to be determined before the algorithm is executed, the iterative equations of the CAS-C algorithm can be described as follows:

$$
\begin{cases}
y_i(t)=y_i\left(t-1\right)^{(1+t^t)} \\
C_{pid}(t)=\left(C_{pid}\left(t-1\right)+\dfrac{7.0}{\psi_{id}}\times V_i\right) \\
\exp\left(\left(1-\exp\left(-\alpha y_i(t)\right)\right)\left(3.0-\psi_{id}\left(C_{pid}(t-1)+\dfrac{7.0}{\psi_{id}}\times V_i\right)\right)\right) \\
-\dfrac{7.0}{\psi_{id}}\times V_i+\exp\left(-2\alpha y_i(t)+b\right) \\
\left(pbest_{pid}\left(t-1\right)-C_{pid}\left(t-1\right)\right)
\end{cases}
\tag{6}
$$

where,

t denotes the number of current iterations and $t-1$ denotes the number of previous iterations;

$C_{pid}(t)$ denotes the current position of the i th ant at the p th target center in d -dimensional space;

$pbest_{pid}(t-1)$ denotes the best position of the i th ant of the p th goal center in the d -dimensional space at step $t-1$.

The following will show the exact order of the algorithm execution in order to demonstrate how CAS-C algorithm is implemented. Assuming that the number of clusters to be obtained is k, then the CAS-C algorithm will be implemented as follows:

1) Initialization. Before the CAS-C algorithm starts iterating, it needs to preset some parameters and assign them certain initial values. The specific settings are as follows: set the search range φ^d, the number of ants n, the maximum number of iteration steps $lstep$, the organization factor r_i, and the organization variable y_i according to the data set. Let $t=1$ and randomly generate n ant locations in the search space for each clustering center.

2) At the beginning of the iteration, each ant moves their location according to Eq. (6). After each iteration, the best position of the i th ant is calculated according to Eq. (5) and recorded for the next iteration. The Euclidean distance is chosen in the calculation to measure the distance of each ant in the data space.

3) Let $t=t+1$ and go to step 2. The algorithm terminates when the number of execution steps reaches the preset maximum number of iteration steps $lstep$ and goes to step 4.

4) Marking the clustering center. After the termination of the iteration, the algorithm converges to a number of points in the space, i.e., all the ants move to a few fixed positions in the data space, which are the cluster centers finally obtained by the clustering algorithm.

5) Split the data and then obtain the clusterings. As per the derived clustering center, as per equation (4), every data in the dataset is assigned to its respective category and the ultimate

clustering output is acquired.

This algorithm is a clustering algorithm based on intelligent optimization, which also has all the advantages of this class of clustering algorithms, mainly in the following three aspects: insensitive to the initial value of the center, able to find the global optimal solution; applicable to the dataset containing classes with multiple sizes and densities; applicable to high-dimensional data.

2.3 Association Rule Mining and Marketing Recommendations

2.3.1 Construction of travel route association rules for online travel users

The user behavior data of the smart tourism platform indicates, on one side, the properties of the online tourism users and the level of activity and hierarchy of online tourism users network behavior (searching tourism information, booking tourism products, sharing and interacting with the network, etc.) prior to, during, and after the tour, and the growing popularity of online tourism due to the high pace of Internet and big data technology development, and on the other side, the questionnaire survey is hardly able to get the data that does not have a common pattern in the course of the tour, thus the data is incomplete. Conversely, the questionnaire survey cannot be used to acquire the lack of travel route data during the entire tour, which results in the incompleteness of the data on various forms of network behaviors of online tourism users and isolates them into islands of data. Thus, there is a significant implication of unlocking the value of user behavior data of smart tourism platform through web crawler technology acquisition of additional user behavior data of smart tourism platform and building association rules on the basis of the two concepts, i.e., the concept of user and tourism elements (destination attraction and route chains), which are of great significance in the development of big data. Unlocking the value of user behavioral data of smart tourism platform and achieving the true potential of the data of online tourism users has a far-reaching significance in the era of big data. In the case of the behavioral data following clustering, this chapter will use Apriori association algorithm to determine the relevance of tourist attractions in the same region of tourism users to enable the generation of the most preferred tourism routes in the same region among similar users.

In this chapter, the travelogues shared by users after traveling are selected as the source data for the study, and the travel route data in the travelogues are extracted as the study data. In order to select a suitable platform for collecting travel itinerary information, combined with the interview data of online travel users, the platforms highlighted were compared, and finally the user travel itinerary information of the A travel website platform was selected as the collection platform. In view of the availability of data, this paper selects the data of tourism nodes as the subject of correlation analysis, and selects N city for empirical research.

To guarantee that the data is authentic and valid, screening conditions of data of tourism route nodes are established:

To begin with, incomplete or garbled user attributes and tourism data in the web crawler are removed;

Secondly, visiting users are not residents of the locality (excluding users who have spent less than or equal to 1.0 days touring).

Thirdly, they have paid a visit to at least one tourist attraction at the destination. The above filtering criteria will yield 1500 travel route data that satisfy the condition of association mining. To include in the input some scattered nodes to be merged as tourism nodes, the data that meet the requirements of association mining are to be considered. the specific method is: for the tourism route data in the database to establish a data dictionary, and statistical word frequency, calculate the cosine similarity between the names of attractions that constitute the tourism routes, and select the names of attractions with cosine similarity ≦ 0.50, and update the names of attractions.

2.3.2 Forms of expression of association rules

Let $I = \{i_1, i_2, ..., i_m\}$ represent the universe of data items. The task-specific dataset is treated as a set of database transactions, where each transaction T is a subset of items such that $T \subseteq I$ .Every transaction is assigned a unique identifier, referred to as a TID. For an itemset A, transaction T is said to contain A if and only if $A \subseteq I$, An association rule is an implication of the form“$A \subseteq B$”, where $A \subset I$, $B \subseteq I$, and $A \cap B = \varnothing$. A rule $A \subseteq B$ holds within the transaction dataset D with support s and confidence Specifically, support s means that s% of transactions in D contain all items in A∪B, while confidence c indicates that c% of transactions in D that contain A also contain B These metrics are defined formally below:

$$\sup\left(A \Rightarrow B\right) = P\left(A \cup B\right) \tag{7}$$

$$conf\left(A \Rightarrow B\right) = P\left(B \cap A\right) \tag{8}$$

A strong association rule is defined as one that meets both the minimum support and minimum confidence thresholds. For convenience, these thresholds are commonly abbreviated as $\min\_\sup$ (minimum support) and $\min\_\mathrm{conf}$ (minimum confidence), both of which range from 0.00% to 100.00%. A group of data items is referred to as an itemset, with a k-itemset. denoting an itemset containing exactly k items. For example, the set {computer, financial_management_software} constitutes a 2-itemset. The occurrence frequency of an itemset, also known as its support count, is the total number of transactions in dataset D that contain all items in the set. When this frequency meets or exceeds the minimum support threshold, the itemset is classified as a frequent k-itemsets. , and the set of all such frequent itemsets is denoted as L^k .Association rule mining proceeds in two key phases: first, identifying all frequent itemsets, which by definition appear at least as often as the predefined minimum support count; second, generating strong association rules from these frequent itemsets, ensuring each rule satisfies both the minimum confidence and minimum support criteria.

2.3.3 Basic idea of association rule algorithm

The basic idea of the current algorithm for discovering large itemsets consists of the following steps:

Test the support of itemsets of length 1.0, called 1-itemsets, by scanning the database. Discard the itemsets that do not fulfill the requirement of minimum support.

Each time an item is added, expand the large 1-item set into a 2-item set, generate all candidate item sets of length 2-, check all candidate item sets by scanning the database, and discard those 2-item sets that do not satisfy the minimum support.

Repeat the above steps: at step K, the $(K-1)$-itemsets found earlier are expanded to K-itemsets and tested to be satisfying minimum support.

Repeat these steps until a larger itemset is found.

2.3.4 Tourist attraction recommendation based on association rule extraction

The key step in a recommendation algorithm based on user group access patterns is to extract core attractions from association rules that can be ranked in terms of their importance. In this paper, we will first summarize the core steps, and then introduce and briefly analyze the key code of this recommendation scheme.

The core steps of the recommendation algorithm based on user group visit patterns can be summarized as follows:

1) Read in the set of attraction association rules $\text{RuleSet} = \{\text{Condition} --> \text{Result}\}$, where Condition represents the conditions of the association rules and Result represents the results of the association rules;

2) For each association rule, combine the attractions in condition $Condition = [ConAttraction_1, ConAttraction_2, ConAttraction_3, \cdots]$ and result $\mathrm{Re}\,sult = [\mathrm{Re}\,sAttraction_1, \mathrm{Re}\,sAttraction_2, \mathrm{Re}\,sAttraction_3, \cdots]$ to generate attraction pair set $AttSet = \{ConAttraction_i, \mathrm{Re}\,sAttraction_j\}(i < Condition.length, j < \mathrm{Re}\,sult.length)$;

3) According to the set of pairs of attractions $AttSet$, statistically calculate the weight of each attraction in the relationship network, and generate a list of recommended access patterns for user groups of the intelligent tourism platform in descending order of weight.

3 Data Mining and Marketing Recommendation Practice

3.1 Data Preprocessing and Cluster Analysis

3.1.1 Distribution of Attribute Values for User Behavior Data

The cluster analysis algorithm needs to have strict demands concerning the representation of attributes value of the data that should be classified; therefore, at the stage of preprocessing the data, the obtained heterogeneous data of the user behavior of the smart tourism platform are made homogeneous by transforming them into numeric attribute values and replacing each attribute value with a distinct number, including 1, 2, 3, etc. Figure 2 shows the results of the distribution of some user attribute values after preprocessing. The preprocessing of the data includes the selection of eight attributes such as gender, age, platform usage time, travel time, number of trips, travel mode, travel destination, number of travelers, etc., data cleaning, integration, discretization, etc., and the numerical conversion of the heterogeneous data is finally completed.

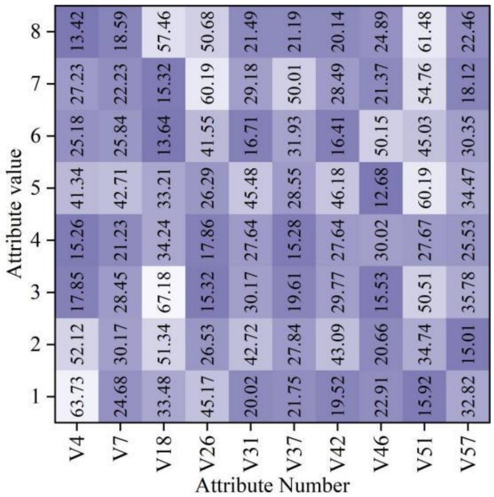

Figure 2: Distribution of user attribute values(Parts)

3.1.2 Analysis of CAS-C clustering results

The global clustering optimization is performed in combination with CAS-C algorithm to avoid the situation of insufficient data clustering accuracy. Table 1 shows the final clustering results. The users are divided into 6 classes according to the behaviors of the smart tourism platform, and each class contains 2 parts of tourism behavior description and demographic description. The behavioral characteristics of users in each class have obvious differences. For example, in Class 1, the number of times users traveled is mostly 1~2 times (56.48%), they are more inclined to choose to travel with a group (52.61%), the number of travel days is around 4~8 days (64.38%), and most of them are males (51.03%) under the age of 35 (56.39%) with postgraduate education or above (50.62%).

Table 1: CAS-C clustering results(Including some data features)

	Category						Average value
	C1	C2	C3	C4	C5	C6	
Description of Travel Behavior							
V4 Number of trips							
1~2	56.48	60.29	59.22	62.37	52.01	65.46	59.31
3~6	24.67	20.35	40.21	29.05	38.22	33.15	30.94
>6	18.85	19.36	0.57	8.58	9.77	1.39	9.75
V18 Mode of travel							
Independent travel	47.39	50.13	42.01	56.72	51.26	54.28	50.30
Traveling in a group	52.61	49.87	57.99	43.28	48.74	45.72	49.70
V26 Number of travel days							
≦3	4.05	24.28	10.03	36.47	12.84	17.37	17.51
4~8	64.38	58.36	42.84	50.12	80.26	64.92	60.15
≧9	31.57	17.36	47.13	13.41	6.90	17.71	22.35
Demographic description							
V37 Gender							
Male	51.03	40.38	49.72	51.36	55.27	46.18	48.99
Female	48.97	59.62	50.28	48.64	44.73	53.82	51.01
V42 Age							
≦35	56.39	50.14	63.02	45.49	55.21	59.86	55.02
>35	43.61	49.86	36.98	54.51	44.79	40.14	44.98
V51 Educational level							
Undergraduate and below level	49.38	41.02	50.37	51.28	60.22	49.87	50.36
Master's degree or above	50.62	58.98	49.63	48.72	39.78	50.13	49.64

3.1.3 Visualization of clustering results

The clustering results are visualized to make it easy to visualize the classification clarity of each cluster. Figure 3 shows the visualization of the clustering results of CAS-C algorithm. From the visualization of the clustering of the 6 categories, there is a certain distance between the horizontal and vertical coordinates of each category, and there is no overlapping area, which proves that the clustering effect is better, and it can effectively distinguish different behavioral data. From the judgment rate of each clustering, the classification accuracy of the six categories for the user behavior data of smart tourism platform reaches 93.69%~96.42%, and the clustering effect is good.

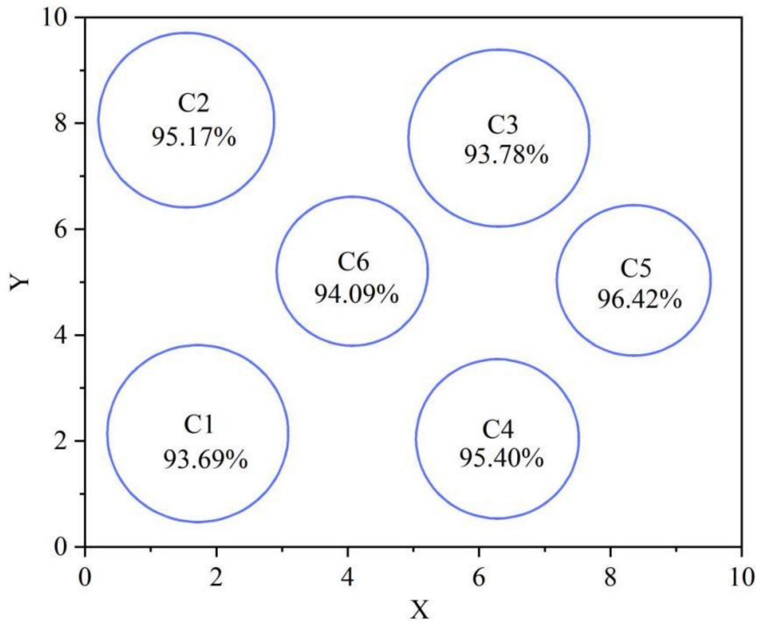

Figure 3: Visualization of the clustering results of the CAS-C algorithm

3.2 Analysis of association rule mining results

Having performed the clustering, the association rule algorithm is applied to extract the association rules between user behavior and core attractions and determine if the obtained rules satisfy the conditions of minimum support threshold and the minimum confidence threshold. Table 2 presents the outcomes of the minimum support and minimum confidence computations of the items prior to and after the mined association rules. The first 15 association rules mined have the minimum support in the range of 76.38-85.41 percent and minimum confidence in the range of 79.02-85.29 percent,with a small difference between the support and confidence, and all of them exceed the critical level of 65%. The mined association rules can be used as a reference for the marketing recommendation of attractions and routes on the smart tourism platform.

Table 2: Minimum support and confidence of association rules(Top 15)

Following item	Previous item	Support Rate (%)	Confidence level (%)	Rule support	Improvement
During the trip, search for travel information online	Gender: Male	80.13	81.22	80.13	1.326
Use the navigation function of the smart phone to find places during the trip	Gender: Female	76.38	83.31	76.38	1.381
Make online hotel reservations	Gender: Male	79.04	82.45	79.04	1.305
Make online flight or train ticket reservations	Gender: Female	81.35	81.53	81.35	1.326
Check travel product information online before the trip	Gender: Male	83.62	80.26	83.62	1.337
Use the navigation function of the smart phone to find places during the trip	Education level: Bachelor's degree	85.41	83.51	85.41	1.313
Make online hotel reservations	Education level: Master's degree	79.45	84.08	79.45	1.327
Make online flight or train ticket reservations	Education level: Bachelor's degree	80.26	83.65	80.26	1.186
Check travel product information before the trip	Education level: Master's degree	82.17	79.02	82.17	1.324
Use the navigation function of the smart phone to find places during the trip	Age: >35	80.27	81.27	80.27	1.061
Make online hotel reservations	Age: ≤35	82.73	83.31	82.73	1.272
Make online flight or train ticket reservations	Age: >35	81.09	80.28	81.09	1.175
Before the trip, determine the destination	Age: ≤35	82.32	81.35	82.32	1.262
Browse the recommended attractions before the trip	Education level; Undergraduate degree Age:>35	85.36	84.44	85.36	1.805
Customize the travel route	Educational Background: Postgraduate degree Age:≤35	81.94	85.29	81.94	1.317

3.3 Effectiveness of Route Recommendation under Cluster Analysis and Association Rule Mining

3.3.1 Statistics on the length of visits and number of clicks by platform users

Based on the clustering results of user behavior and the mined association rules, the weights of each core attraction and user interest are calculated, after which a list of core attractions and travel routes recommendations for smart tourism platform users is generated. The visit time of the optimized and updated smart tourism platform users is counted to analyze the user's interest in the recommended results. Figure 4 shows the average duration of users' visits to the smart tourism platform users and the average number of clicks on the recommended attractions/routes in a month. The average duration of users' visit to the smart tourism platform within 30 days is around 40.22min~83.36min, and the average number of clicks on recommended attractions/routes is 16.46~30.76 times. It can be inferred from the data that the recommended attractions and routes of the platform pushed according to the clustering results and association rules are more attractive to the target users.

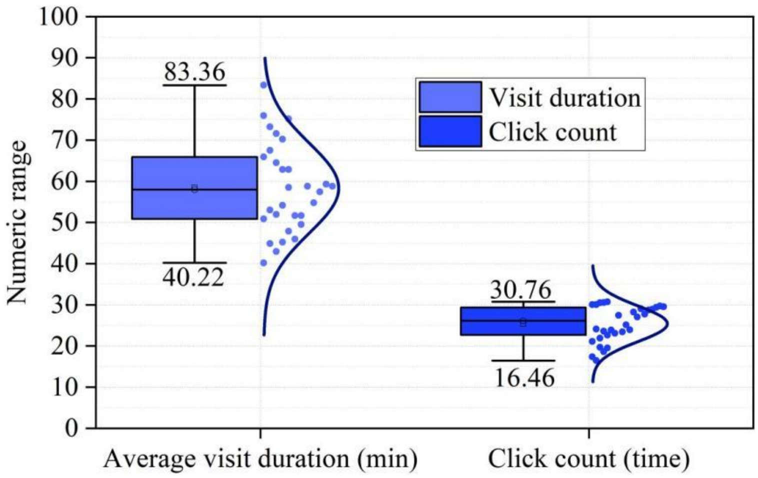

Figure 4: Average duration of users' visits and the average number of clicks

3.3.2 Analysis of the relationship between recommended attractions and user interest level

Both the number of suggested attractions and routes influences recommendation accuracy of the algorithm, as well as the interest level of users in the respective attraction or route. The relationship between the number of attractions being recommended and the degree of interest of the user is depicted in Figure 5. As the suggested attractions reduce slowly to 50, there is a growing tendency in the user interest level and recommendation accuracy.When the algorithm recommends only 3 attractions at a time, the user's interest level reaches the highest 0.75 and the recommendation accuracy reaches 80.60%.

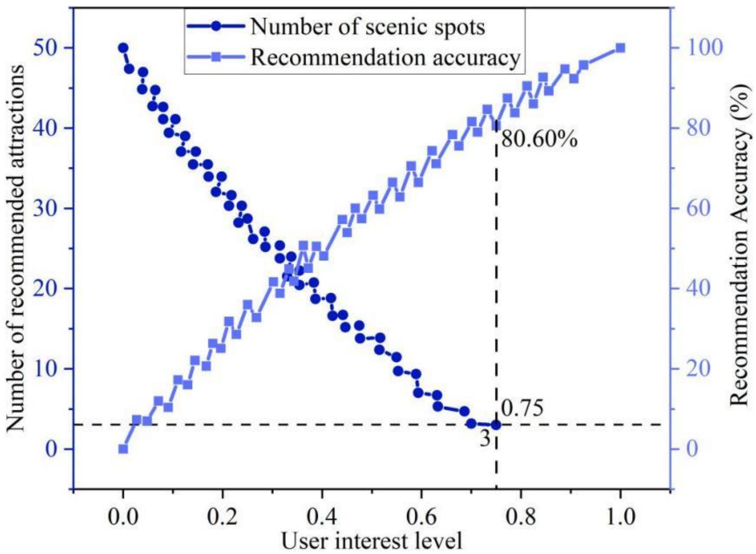

Figure 5: Relationship between the number of attractions and the level of interest

3.3.3 Analysis of the relationship between recommended routes and user interest level

Figure 6 illustrates the correlation between the number of the proposed routes and the interest level of the user. When there is only 1 recommended route, the user's interest degree reaches 1.0 and the recommendation accuracy reaches 100%, which is generally a route tailored for new users by the smart tourism platform. With the increase of recommended routes, the user interest degree and recommendation accuracy also show a decreasing trend. However, considering the marketing cost of the smart tourism platform, it is difficult to cover the operation cost by recommending only 3 attractions or 1 route at a time, so the actual recommendation will give up part of the accuracy.

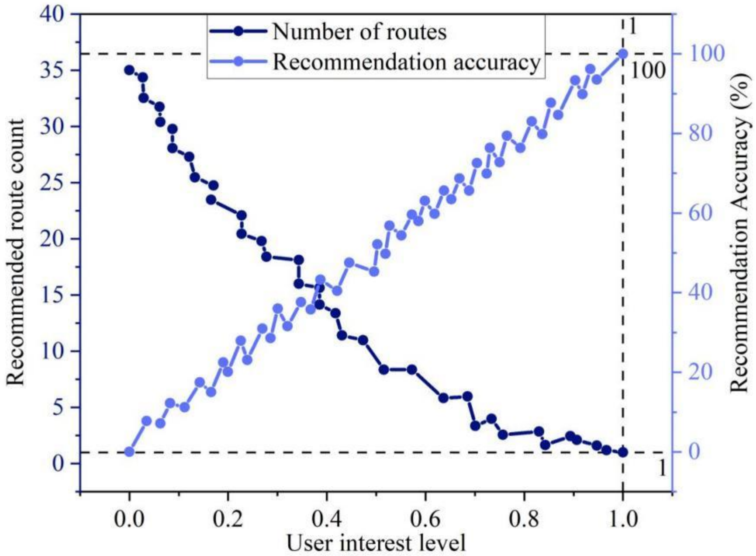

Figure 6:The relationship between the number of routes and the level of interest

3.3.4 Recommendation effects of different numbers of recommended attractions and routes

To figure out the most appropriate amount of recommended attractions and routes within smart tourism platform, the accuracy, recall, coverage and popularity of the algorithms are calculated when recommending different numbers of attractions and routes. Figure 7 shows the performance comparison of the algorithm with different number of recommendations. Overall, when the single recommendation of attractions and the number of attractions totaled 30, the accuracy rate reached 62.46%, the recall rate reached 16.84%, the coverage rate reached 50.36%, and the popularity reached 10.38%, at which time the recommendation effect was the best. Therefore, the total number of single recommended attractions and routes of the smart tourism platform is set to 30 to get the optimal recommendation effect.

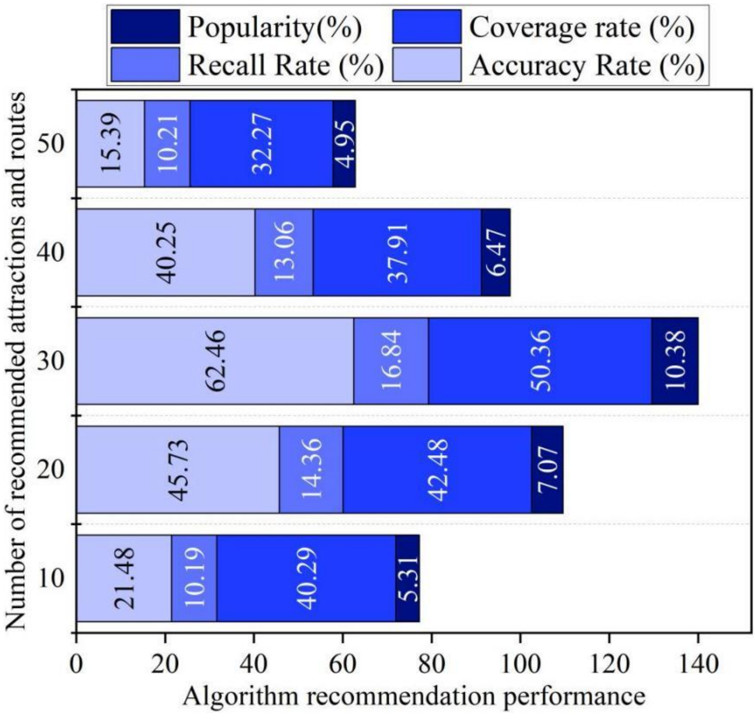

Figure 7: Algorithm performance under different recommendation quantities

4 Conclusion

In this paper, clustering analysis and association rule mining are performed on the user behavior data of the smart tourism platform to calculate the association between the number of recommended attractions/routes and the user's interest level, to realize the precise marketing push for the target users, and to improve the platform stickiness and usage inertia of the target users.

The CAS-C algorithm divides the user behaviors into 6 groups, and the judgment rate of every group is higher than 93%. Mined association rules have a minimum support and minimum confidence of 76.38 - 87.13 and 76.43 - 87.23 respectively. The accurate push determined by the results of the research and analysis increased the time of accessing the platform by the users to 40.22~83.36min and increased the number of clicks on recommended content to 16.46~30.76 times. Combined with the consideration of marketing costs, the total number of single push attractions/routes was finally determined to be 30, achieving a balance between user interest, recommendation accuracy, and marketing costs.

Funding

The article is a research output of the 2024 Henan Provincial Higher Education Teaching

Reform Research and Practice Project entitled as, The Study on the Innovation of the Practical Teaching of the E-commerce Specialization Through the Lens of the Concept of College-Academy Collaborative Education (Project Number: 2024SJGLX0905).

The present article is the outcome of the phased implementation of the key program of the National Social Science Foundation of China in 2021, i.e. Research on Marxist Ideology of People's Democracy (Project No. 21&ZD007) and the research task of the Marxism Theory Research and Construction Program, People's Democracy as the Correct Interpretation of Comprehensively Building a Socialist Modernized State (Project No. 2023MYB010).

References

[1] Sigala, M., Rahimi, R., & Thelwall, M. (2019). Big data and innovation in tourism, travel, and hospitality. Culinary Science & Hospitality Research, 23(6), 1-11.

[2] Theodorakopoulos, L., & Theodoropoulou, A. (2024). Leveraging big data analytics for understanding consumer behavior in digital marketing: A systematic review. Human Behavior and Emerging Technologies, 2024(1), 3641502.

[3] Del Vecchio, P., Mele, G., Ndou, V., & Secundo, G. (2018). Creating value from social big data: Implications for smart tourism destinations. Information processing & management, 54(5), 847-860.

[4] Gu, J. (2022). Research on precision marketing strategy and personalized recommendation method based on big data drive. Wireless Communications and Mobile Computing, 2022(1), 6751413.

[5] Krabokoukis, T. (2025). Bridging neuromarketing and data analytics in tourism: An adaptive digital marketing framework for hotels and destinations. Tourism and Hospitality, 6(1), 12.

[6] Zhou, D. (2024). Tourism destination marketing strategy driven by big data. Journal of Computational Methods in Science and Engineering, 24(4-5), 2593-2609.

[7] Awad, A. (2025). The role of digital marketing tools in promoting tourism: An applied study on online marketing strategies. Innovative Marketing.

[8] Gurung, G., & Upadhaya, A. K. (2024). Analyzing Pattern Recognition Approaches for Forecasting Tourism Demand. Available at SSRN 4904211.

[9] Tang, L. (2025). Deep learning-driven tourist behavior prediction and intelligent recommendation system for smart tourism platforms. Service Oriented Computing and Applications, 1-17.

[10] Song, M. (2025). Research on intelligent decision support platform for tourism enterprises based on multi-source heterogeneous data fusion. Scientific Reports, 15(1), 39810.

[11] Pan, B., & Yang, Y. (2017). Monitoring and forecasting tourist activities with big data. Management science in hospitality and tourism: Theory, practice, and applications, 43-62.

[12] Li, D., Deng, L., & Cai, Z. (2020). Statistical analysis of tourist flow in tourist spots based on big data platform and DA-HKRVM algorithms. Personal and Ubiquitous Computing, 24(1), 87-101.

[13] Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A big data analytics method for tourist behaviour analysis. Information & Management, 54(6), 771-785.

[14] Zhang, J., & Gao, M. (2025). The application of deep learning in economic analysis and marketing strategy formulation in the tourism industry. PLoS One, 20(6), e0321992.

[15] Devkota, N., Gajdka, K., Siwakoti, R., Klímová, M., & Dhakal, K. (2023). Promoting sustainable tourist behavior through promotional marketing. Journal of Tourism and Services, 14(26), 219-241.

[16] Huang, X., & Wei, Z. (2024, December). Research on User Behavior Analysis and Precision Marketing Strategy of Smart Tourism. In 2024 4th International Conference on Social Development and Media Communication (SDMC 2024) (pp. 237-243). Atlantis Press.

[17] Zhang, F., & Li, J. (2024, December). Design of Individualized Tourism Service Recommendation Algorithm Based on User Behavior Analysis. In 2024 International Conference on Internet of Things, Robotics and Distributed Computing (ICIRDC) (pp. 435-441). IEEE.

[18] Zhang, J., Wu, T., & Fan, Z. (2019). Research on precision marketing model of tourism industry based on user's mobile behavior trajectory. Mobile Information Systems, 2019(1), 6560848.

[19] Sakas, D. P., Reklitis, D. P., Terzi, M. C., & Vassilakis, C. (2022). Multichannel digital marketing optimizations through big data analytics in the tourism and hospitality industry. Journal of Theoretical and Applied Electronic Commerce Research, 17(4), 1383-1408.

[20] Gan, M., & Ouyang, Y. (2022). Study on tourism consumer behavior characteristics based on big data analysis. Frontiers in psychology, 13, 876993.

[21] Lacárcel, F. J. S. (2022). Main uses of artificial intelligence in digital marketing strategies linked to tourism. Journal of Tourism, Sustainability and Well-being, 10(3), 215-226.

[22] Selvakumar, P. (2025). Big Data Analytics in Tourism Development and Marketing. In Redefining Tourism With AI and the Metaverse (pp. 249-278). IGI Global Scientific Publishing.

[23] Wu, D., & Ma, Q. (2025). AI-assisted customer behavior analysis and hotel loyalty strategy optimization. Future Technology, 4(3), 97-106.