EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Design Method for Travel E-commerce Platform Based on HHO improved K-means Clustering Algorithm

Mihua Dang[1,*] and Suiming Yang[1]

[1] School of Humanities and Management Xi'an Traffic Engineering Institute, Xi'an 710300, Shaanxi, China

Abstract

Convenient and intelligent tourism product recommendation method, as the key technology of tourism E-commerce platform design, not only provides academic value to the research of tourism E-commerce platform, but also improves the efficiency of personalized recommendation of tourism products. In order to improve the quality of tourism recommendation, this paper proposes a tourism E-commerce platform design method based on HHO improved K-means clustering algorithm. Firstly, the Harris optimization algorithm is used to improve the K-means algorithm to construct a user-oriented tourism product recommendation strategy; then, combined with the XGBoost algorithm, an item-oriented tourism product recommendation strategy is proposed; secondly, the two strategies are mixed to construct a personalized tourism product recommendation model. Finally, the effectiveness of the proposed method is verified by simulation experiment analysis. The results show that the recommendation accuracy of the tourism E-commerce platform design method proposed in this paper reaches more than 90%, and the recommendation response time meets the real-time requirements, which can provide personalized tourism product recommendation for platform users and enhance the purchase of tourism products.

Keywords: travel E-commerce platform design, K-means clustering algorithm, Harris Hawk optimization algorithm, XGBoost

Received on 29 December 2023, accepted on 22 April 2024, published on 26 April 2024

doi: 10.4108/eetsis.5782

*Corresponding author. Email: 443721376@qq.com

1 Introduction

With the rapid growth of the national economic level, people's quality of life and living standards have improved, and tourism consumption accounts for an increasing proportion of people's daily consumption [1]. With the rise of the E-commerce industry, the development of data mining technology, machine learning algorithms, the rapid development of intelligent tourism service technology, tourism E-commerce has become an important means for people to obtain tourism information and tourism product booking [2]. With the further improvement of people's quality of life and the further expansion of tourism demand, most of the tourism E-commerce platform tourism product retrieval is complex, the process is redundant, and the product recommendation interface is uniform, which is difficult to meet the needs of people's tourism consumption [3]. Convenient and intelligent tourism product recommendation method as the key technology of tourism E-commerce platform design, not only provides academic value to the research of tourism E-commerce platform, but also improves the efficiency of personalized recommendation of tourism products [4]. Therefore, the tourism product recommendation method based on intelligent algorithm has important theoretical value and obvious practical application value. The key technology of the tourism E-commerce platform is the study of tourism product recommendation algorithm, which mainly refers to the intelligent analysis of the tourism E-commerce platform user-related information data, the user's consumption potential and consumption tendency prediction and judgment, so as to provide the user with the tourism E-commerce platform with the function of targeting services [5]. Data mining and tourism E-commerce combination of research methods are more, specifically divided into clustering analysis [6], classification analysis [7], correlation analysis [8]. Literature [9] proposed a data classification method based on lightGBM algorithm and applied it to the E-commerce advertisement recommendation problem to achieve personalized advertisement recommendation based on E-commerce platform; Literature [10] proposed an E-commerce platform design method based on fuzzy clustering algorithm, which clustered the E-commerce platform users to provide an auxiliary method for the E-commerce platform users; Literature [11] used the Apriori association rule algorithm to classify E-commerce platform users and provide accurate customer service for E-commerce platform users; Literature [12] uses K-means, PCA and integrated learning algorithms to construct a personalized tourism product recommendation method, and adopts user behavioral data to validate and analyze the proposed method; Literature [13] adopts a deep learning method to identify tourism products analysis, and combined with the evaluation degree of tourism products to give appropriate tourism products; literature [14] constructs a tourism E-commerce platform by improving the K-means clustering algorithm to achieve personalized customization of tourism products. According to the analysis of the above research, although the integration of machine learning algorithms, data mining technology and tourism E-commerce has achieved certain results, there are still some limitations by then [15]:

1) More research on tourism E-commerce design, but the lack of quantitative analysis;

2) Although the tourism E-commerce platform design methodology includes clustering, identification and correlation analysis, it is more studied individually, and lacks the integration of the research;

3) The optimization process of K-means algorithm for personalized recommendation method of tourism products is easy to fall into local optimum, resulting in poor recommendation effect.

Aiming at the problems existing in the current design method of tourism E-commerce platform, this paper proposes a personalized recommendation of tourism products based on the optimization algorithm to improve the K-means clustering optimization algorithm with integrated machine learning algorithm. This paper uses bio-heuristic optimization algorithm to optimize and improve K-means clustering algorithm, constructs tourism product recommendation algorithm, and applies it to tourism E-commerce service. The improved effectiveness and performance enhancement of this paper's method is verified through data analysis.

2 Travel E-commerce platform design analysis

With the development of the Internet and information technology, all kinds of websites provide users with rich and diverse tourism products. Facing the massive amount of information, it is difficult for consumers to obtain the information they are interested in, thus generating the information overload problem [16]. In order to solve the information overload problem, an effective way - personalized recommendation technology has been proposed. Common personalized recommendation techniques include content-based recommendation, demographic-based recommendation, and collaborative filtering recommendation [17]. Collaborative filtering recommendation algorithms include user-based collaborative filtering algorithms and product-based collaborative filtering algorithms [18].

In the tourism E-commerce platform, the tourism product personalized recommendation algorithm needs to satisfy the following three points [19]: 1) the tourism product recommendation result needs to be accurate; 2) the product recommendation needs to be real-time; and 3) the tourism products have diversity.

In order to better achieve rate yo product personalized recommendation, this paper proposes a hybrid collaborative filtering method for tourism products using user-based collaborative filtering algorithm and product-based collaborative filtering algorithm, and the specific design scheme is shown in **Figure 1**.

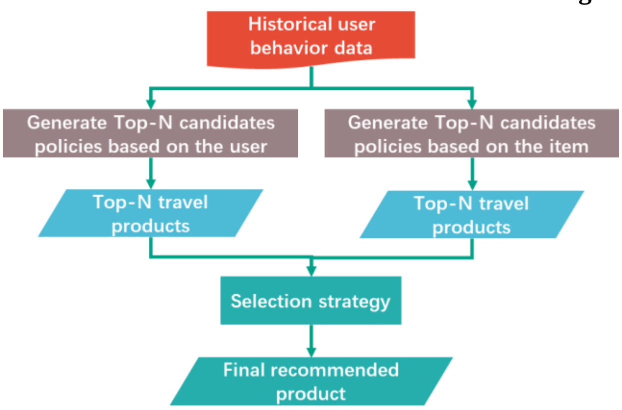

Figure 1 Hybrid collaborative filtering algorithm for tourism products

(1) User Top-N candidate generation strategy based on clustering algorithm

The user-based Top-N candidate generation strategy employs a clustering algorithm to generate candidate items [20] in the following steps:

1) Construct a user-tourism product scoring matrix by scoring tourism products through users' historical behavioral data;

2) Determine the clustering center and use the clustering algorithm to get the clustering results;

3) Pearson correlation similarity is used to calculate user similarity and identify the set of clusters that have the highest similarity to the target user;

4) Calculate user ratings for uncollected travel products and sort the ratings;

5) Generate recommendations.

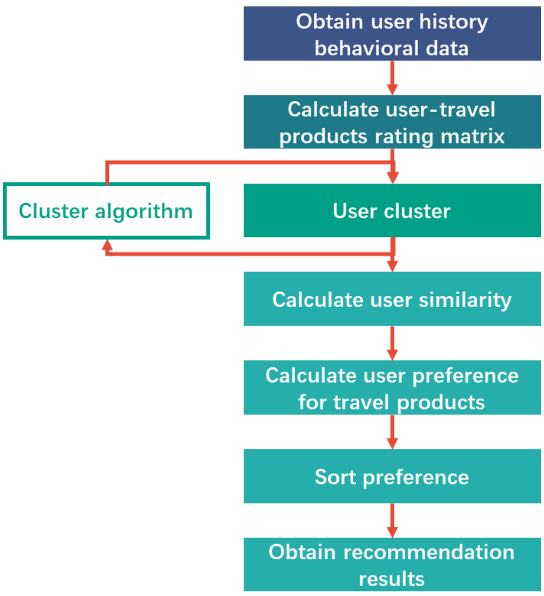

Figure 2 Tourism product recommendation algorithm based on clustering method

(2) Top-M candidate generation strategy for products based on modeling strategy

The product-based Top-N candidate generation strategy uses a classification algorithm to generate candidate items [21] in the following steps:

1) Clean the text data of reviews of tourism products to remove duplicate data and remove missing data;

2) Extract the weak emotion labeling information and construct the input feature set;

3) Using classification algorithm to construct tourism product emotion recognition model;

4) Calculate the similarity between the user requirement label and each travel product label, and rank the ratings;

5) Generate recommendations.

3 Related Technologies

3.1 K-means clustering algorithm

The K-means algorithm is the most popular divisive clustering algorithm, which performs well in handling large data classification [22]. The algorithm determines the clustering centers and the elements to which they belong by minimizing an objective function based on squared error. The aim is to keep the cluster centers as far away from each other as possible and associate each data point to the nearest cluster center. In the K-means algorithm, the Euclidean distance is commonly used as a similarity measure. The objective function of the K-means algorithm is defined as equation (1):

$$J=\sum_{i=1}^{K}\left(\sum_{k}\left\|x_k-c_i\right\|^2\right) \tag{1}$$

where K is the number of clusters, c_i is the center of the cluster, and x_k is the k th data point in the i th cluster.

The specific process of the algorithm is shown in Figure 3 and the detailed steps are as follows:

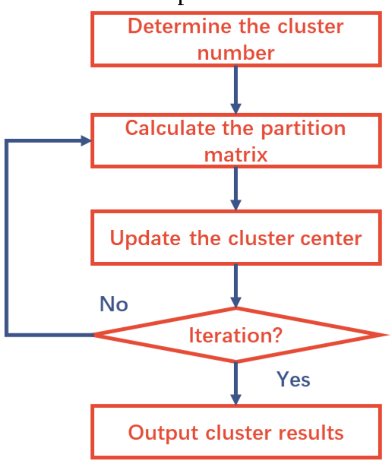

Figure 3 Flowchart of K-means clustering algorithm

Step 1: Determine the total number of classified categories K and randomly select K cluster category centers $C=(c_1,c_1,\cdots,c_K)$.

Step 2: Compute the partition matrix. A data point belongs to the cluster whose center is closest to that data point. Therefore, the clusters are represented by the binary division matrix U . U The elements in are determined as shown in equation (2):

$$u_{ij}=\begin{cases}1 & if\ \left\|x_j-c_i\right\|^2\le\left\|x_j-c_t\right\|^2, \forall t\ne i\\ 0 & otherwise\end{cases} \tag{2}$$

where u_{ij} indicates whether the j th data point belongs to the i th cluster class.

Step 3: Update the cluster centers. Minimize the objective function for each cluster class center c_i Define equation (3):

$$c_i = \frac{\sum_{j=1}^{N} u_{ij} x_j}{\sum_{j=1}^{N} u_{ij}} \quad (3)$$

where N denotes the number of samples.

Step 4: Compute the objective function using equation (1). Verify that the function converges or that the difference between two neighboring values of the objective function is less than a given threshold and stop. Otherwise repeat step 2.

3.2 Harris Hawk Optimization Algorithm

Harris Hawks Optimization (HHO) [23] is a meta-heuristic algorithm inspired by the collaborative foraging behavior of Harris Hawks. Harris hawks are raptors found in southern Arizona, USA, that efficiently perform collaborative foraging in several phases including tracking, circling, and attacking.

(1) Initialization of stocks

The HHO population was initialized using a random uniform distribution strategy, Eq. (4):

$$X_i = lb + rand \cdot (ub - lb) \quad (4)$$

where X_i denotes the *ith* Harris's hawk individual, and lb and ub denote the upper and lower limits of the search space, respectively.

(2) Calculation of adaptation values

Based on the objective function, the fitness value Eq. (5) is calculated:

$$fitness_i = f(X_i) \quad (5)$$

where $fitness_i$ denotes the value of the adaptation of the *ith* individual Harris's hawk and $f(\cdot)$ denotes the objective function.

(3) Calculate the escape energy factor

In order to balance the exploration and development capabilities of the HHO algorithm, the Harris Hawk's shift from global to local search relies mainly on the escape energy factor E to control the change of values as shown in Figure 4, which is calculated as follows:

$$E = 2E_0(1 - t/T) \quad (6)$$

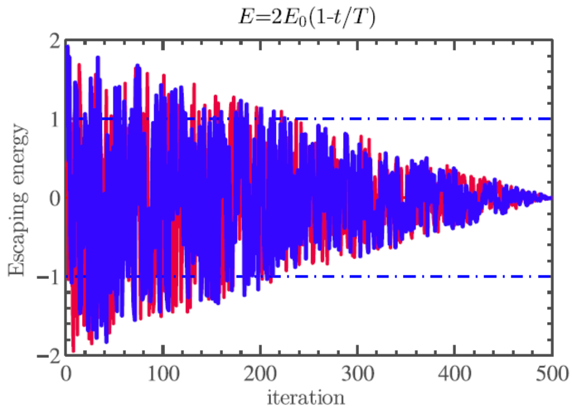

Figure 4 Variation of fugitive energy factor

Where E_0 is a random number from -1 to 1, t is the current iteration number, and T is the maximum iteration number.

(4) Exploration phase

When $|E| > 1$, the HHO algorithm enters the exploration phase, where the Harris Hawk randomly perches at a number of locations, tracks and detects prey through its keen eyes, and hunts with two equal-opportunity strategies, Eq. (7) and Eq. (8):

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1 \cdot |X_{rand}(t) - 2r_2 X(t)| & q \geq 0.5 = \\ (X_{rabbit}(t) - X_m(t)) - r_3 \cdot (lb + r_4 \cdot (ub - lb)) & q < 0.5 \end{cases} \quad (7)$$

$$X_m(t) = \frac{1}{N} \sum_{i=1}^{N} X_i(t) \quad (8)$$

Among them, $X_{rand}(t)$ is the randomly selected individual in the current population, $X_{rabbit}(t)$ is the current optimal individual, $X_m(t)$ is the average position of the current population, r_1 , r_2 , r_3 , r_4 are random numbers from 0 to 1, ub and lb are the upper and lower bounds of the population, and N is the number of the population.

(3) Development phase

After locating the target prey, the Harris Hawk will form a circle around the prey and wait for an opportunity for a surprise attack. The actual hunting process is complex, and the Harris Hawk can make the necessary adjustments to round up prey based on the current state. In order to better grok your hunting behavior, the development phase is updated using four strategies and a parameter E and a random number from 0 to 1 to decide which strategy to use.

1) Soft Surrounding

When $|E| \geq 0.5$ and $r \geq 0.5$, the prey has enough energy to try to escape from the encirclement by random jumps, but ultimately cannot, so the Harris Hawk hunts using a soft encirclement approach, Eq. (9)-Eq. (11):

$$X(t+1) = \Delta X(t) - E\left|JX_{rabbit}(t) - X(t)\right| \quad (9)$$

$$\Delta X(t) = X_{rabbit}(t) - X(t) \quad (10)$$

$$J = 2(1 - r_5) \quad (11)$$

Where $\Delta X(t)$ is the difference between the optimal individual and the current individual, r_5 is a random number uniformly distributed from 0 to 1, and J is the hopping distance of the rabbit during the escape.

2) Hard Surround

When $|E| < 0.5$ and $r \geq 0.5$, the prey has neither enough energy to get away nor the chance to escape, so the Harris Hawk hunts using the hard encirclement approach model, and the hard encirclement strategy is shown in Figure 5:

$$X(t+1) = X_{rabbit}(t) - E\left|\Delta X(t)\right| \quad (12)$$

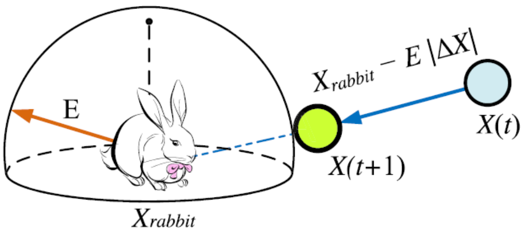

Figure 5 Hard Surround

3) Use a soft envelope of progressive fast swoops

When $|E| \geq 0.5$ and $r < 0.5$ the prey has a chance to escape from the encirclement and the escape energy is sufficient, so the Harris Hawk needs to form a smarter soft encirclement before attacking, which is implemented through the following two strategies as shown in Figure 6. The first strategy is specifically:

$$Y = X_{rabbit}(t) - E \cdot \left|JX_{rabbit}(t) - X(t)\right| \quad (13)$$

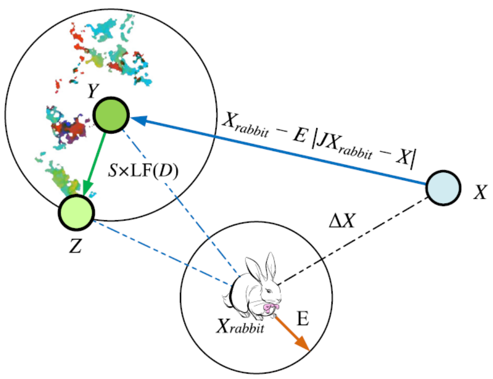

Figure 6 Soft envelopment of a progressive fast dive

When the first strategy is invalid, the second strategy is executed:

$$Z = Y + S \times LF(D) \quad (14)$$

where D is the problem dimension, S is a random variable with D, and LF denotes the Levy flight function, formulated as equation (15):

$$LF(x) = 0.01 \times \frac{l \times m}{|\mu|^{\frac{1}{\beta}}} \quad (15)$$

$$\sigma = \left(\frac{\Gamma(1+\beta) \times \sin(\pi\beta/2)}{\Gamma((1+\beta)/2) \times \beta \times 2^{\frac{\beta-1}{2}}}\right)^{\frac{1}{\beta}} \quad (16)$$

where l and m are uniformly distributed random numbers from 0 to 1, and β is a constant taking the value 1.5. The final calculation of this strategy is shown in equation (17):

$$X(t+1) = \begin{cases} Y & F(Y) < F(X(t)) \\ Z & F(Z) < F(X(t)) \end{cases} \quad (17)$$

4) Hard bracketing using progressive fast dives

When $|E| < 0.5$ and $r < 0.5$, the prey has a chance to escape but the escape energy is not enough, so Harris Hawks form a hard encirclement prior to the raid, reducing the average distance between them and the prey, and hunt using the following strategy, as shown in Figure 7:

$$Y = X_{rabbit}(t) - E \cdot \left|JX_{rabbit}(t) - X_m(t)\right| \quad (18)$$

$$Z = Y + S \times LF(D) \quad (19)$$

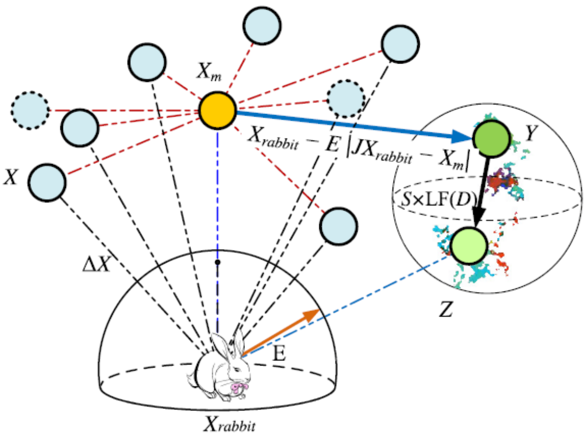

Figure 7 Hard envelope for progressive fast dive

(4) Pseudo-code and flowchart

According to the analysis of the principle and mechanism of HHO algorithm, the stages of the search of HHO algorithm are shown in Figure 8, the specific flow is shown in Figure 9, and the pseudo-code is shown in Figure 10.

Step 1: Initialize the number of HHO populations with the number of iterations;

Step 2: Initialize the HHO population. Initialize the HHO population using the random uniform distribution strategy, calculate the fitness value, and obtain the current optimal value and optimal solution;

Step 3: Calculate the escape energy factor and jump distance.

Step 4: Population position update. When $|E|>1$ is used, the optimization update is performed using the exploratory phase position update strategy; when $|E|\geq 0.5$ and $r\geq 0.5$ are used, the soft encirclement strategy is used; when $|E|\geq 0.5$ and $r<0.5$ are used, the soft encirclement with progressive fast swooping is used; when $|E|<0.5$ and $r\geq 0.5$ are used, the hard encirclement strategy is used; and when $|E|<0.5$ and $r<0.5$ are used, the hard encirclement with progressive fast swooping is used;

Step 5: Calculate the fitness value and update the current optimal value with the optimal solution;

Step 6: Determine whether the number of iterations reaches the maximum number of iterations. If it reaches, output the optimal solution and optimal value; otherwise, return to step 3.

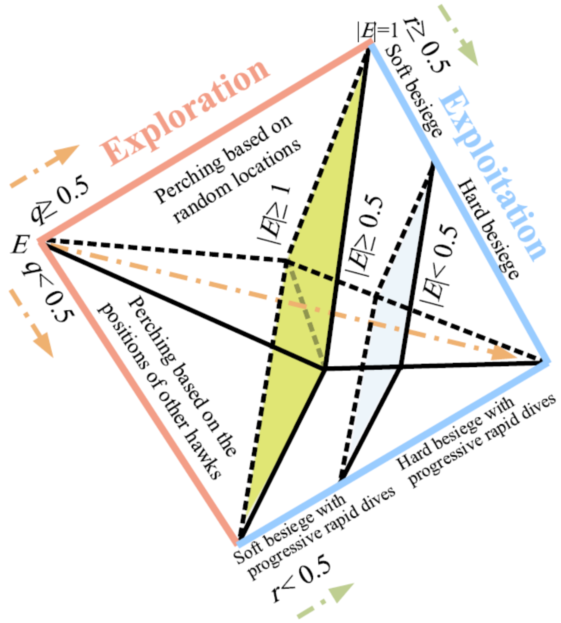

Figure 8 Conversion of HHO algorithm stages

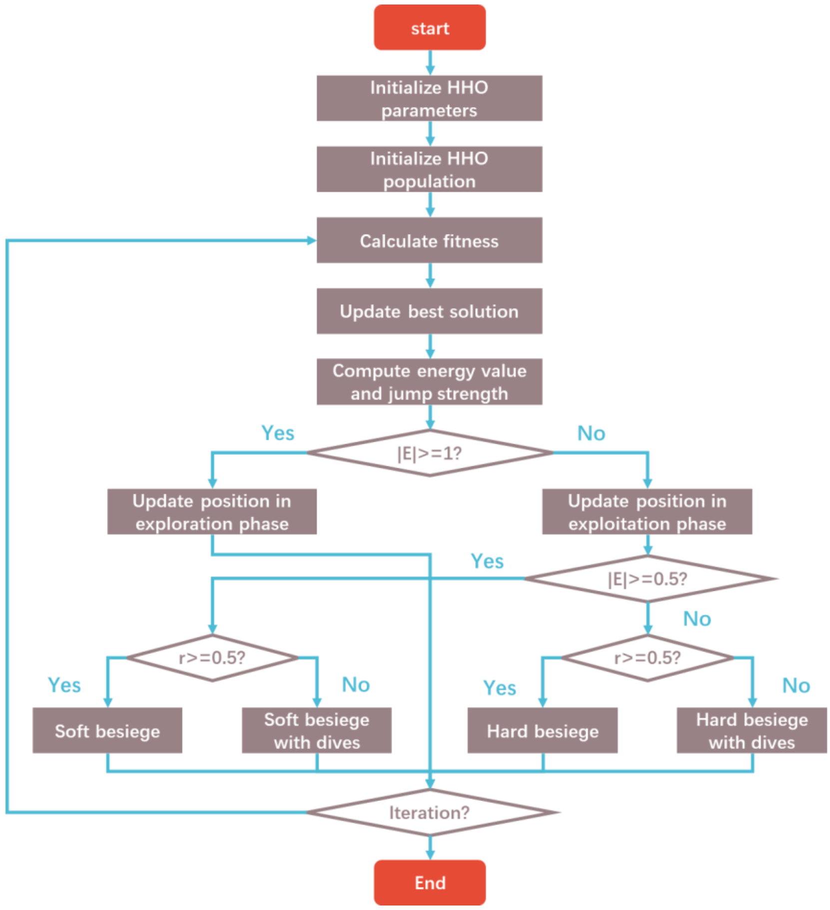

Figure 9 Flowchart of Harris Hawk optimization algorithm

```
Algorithm 1: Pseudo code of HHO algorithm
Assign parameters to the HHO algorithm;
Initialize the population of HHO algorithm;
Calculate fitness of population;
While stopping condition is not met do
    Set best solution;
    Update energy and jump strength;
    if |E|>=1
        Update the position of population using exploration phase;
    else
        if |E|>=0.5
            if r>=0.5
                Update position using soft besiege strategy;
            else
                Update position using soft besiege with progressive rapid dives;
            end
        else
            if r>=0.5
                Update position using hard besiege strategy;
            else
                Update position using hard besiege with progressive rapid dives;
        end
    end
Compute fitness of population;
end
Output best solution.
```

Figure 10 Pseudo-code of the HHO algorithm

3.3 XGBoost Algorithm

(1) Base Classifier - Random Forests

Random Forest (RF) is an integrated learning algorithm based on decision trees, and its algorithmic principle is based on the Bagging integration algorithm and the random subspace method [24].The RF algorithm utilizes the samples to be classified, trains to produce a decision tree, and determines the predicted classification results of all the results of all the decision trees by aggregating the results by voting, and its specific flowchart is shown in Figure 11 shown.

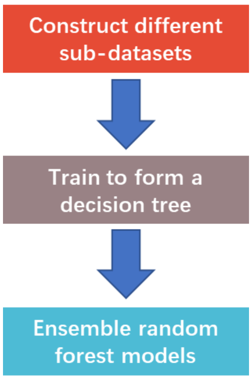

Figure 11 Schematic diagram of a random forest

(2) XGBoost algorithm

XGBoost (Extreme Gradient Boosting) algorithm implements the generation of weak learners by optimizing the structured loss function, and uses strategies such as pre-sorting and weighted quantile to improve the performance of the algorithm [25].The base classifier of XGBoost algorithm is RF algorithm, which is a supervised model based on random forest, and the specific model is as follows:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F \tag{20}$$

where K denotes the number of random forests; x_i denotes the input vector and $x_i \in R^m$; F denotes the random forest function space and f_k denotes a specific random forest in the function space.

The objective function of the XGBoost model is:

$$obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{21}$$

where $\sum_{i=1}^{n} l(y_i, \hat{y}_i)$ is the loss function and $\sum_{k=1}^{K} \Omega(f_k)$ is the regular term, which is mainly used to smooth the weights learned by the model to avoid overfitting phenomenon.

In this paper, the squared error function is used as a function of loss, which is calculated as follows:

$$L\left(y,\hat{y}\right)=\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)^2 \tag{22}$$

The formula for the regular term is as follows:

$$\Omega\left(f_k\right)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}\omega_j^2 \tag{23}$$

where γ and λ are customized parameters and ω_j is the weight of leaf j.

The specific flowchart of the XGBoost algorithm is shown in Figure 12, and the steps of the training process are as follows:

1) During each iteration, keep the existing model and add the newly generated random forest model;

$$\hat{y}_i^{(t)}=\hat{y}_i^{(t-1)}+f_t\left(x_i\right) \tag{24}$$

Where, $\hat{y}_i^{(t)}$ denotes the tth iteration number to get the model, $\hat{y}_i^{(t-1)}$ denotes the t-1st iteration number to get the model, and $f_t\left(x_i\right)$ denotes the tth iteration number to add the random forest model.

2) Use the objective function to determine the merits of the additional random forest model:

$$obj^{(t)}\left(\theta\right)=\sum_{i=1}^{n}l\left(y_i,\hat{y}_i^{(t-1)}+f_t\left(x_i\right)\right)+\Omega\left(f_t\right)+\sum_{i=1}^{t-1}\Omega\left(f_i\right) \tag{25}$$

obtained by Taylor expansion:

$$obj^{(t)}\left(\theta\right)\approx\sum_{i=1}^{n}\left[l\left(y_i,\hat{y}_i^{(t-1)}\right)+g_i f_t\left(x_i\right)+\frac{1}{2}h_i f_t^2\left(x_i\right)\right]+\Omega\left(f_t\right)+\sum_{i=1}^{t-1}\Omega\left(f_i\right) \tag{26}$$

Where $g_i=\partial_{\hat{y}_i^{(t-1)}}l\left(y_i,\hat{y}_i^{(t-1)}\right)$ is the first order gradient of the loss function and $h_i=\partial_{\hat{y}_i^{(t-1)}}^2 l\left(y_i,\hat{y}_i^{(t-1)}\right)$ is the second order gradient of the loss function. Since $l\left(y_i,\hat{y}_i^{(t-1)}\right)$ is a constant and $\sum_{i=1}^{t-1}\Omega\left(f_i\right)$ is a constant, the above equation can be simplified by removing the constant term:

$$obj^{(t)}\left(\theta\right)\approx\sum_{i=1}^{n}\left[g_i f_t\left(x_i\right)+\frac{1}{2}h_i f_t^2\left(x_i\right)\right]+\Omega\left(f_t\right) \tag{27}$$

According to the basis function principle, the objective function can be transformed into:

$$obj^{(t)}\left(\theta\right)\approx\sum_{j=1}^{T}\left[G_j\omega_j+\frac{1}{2}\left(H_j+\lambda\right)\omega_j^2\right]+\gamma T \tag{28}$$

where $G_j=\sum_{i\in I_j}g_j$, $H_j=\sum_{i\in I_j}h_j$. The weights are obtained when the objective function is optimal by the derivation of ω_j as follows:

$$\omega_j^*=-\frac{G_j}{H_j+\lambda} \tag{29}$$

The optimal loss function can be obtained from the above equation:

$$obj^{(t)}=-\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j+\lambda}+\gamma T \tag{30}$$

Based on the optimal objective function, the optimal random forest is determined, added to the XGBoost model, and iterated.

3) Enumerate different random forest structures using greedy algorithm.

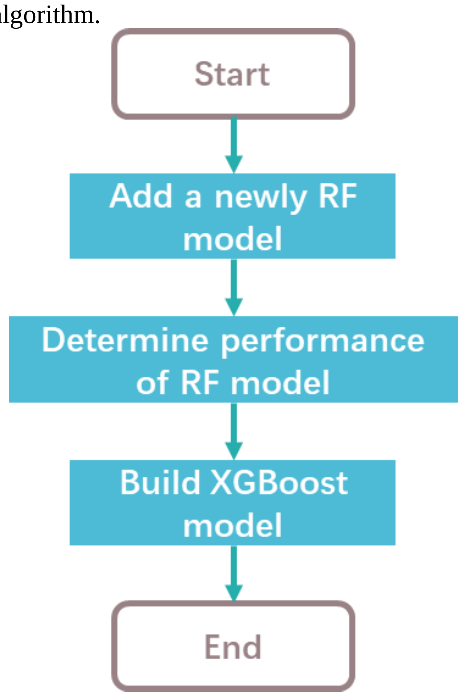

Figure 12 XGBoost flowchart

4 Personalized Tourism Product Recommendation Model Based on HHO Algorithm Optimized K-means and XGBoost Algorithm

(1) Personalized Tourism Product Recommendation Model Based on Improved K-means Clustering Method of HHO Algorithm

In order to increase the accuracy of the K-means clustering method, this paper uses the HHO algorithm in combination with K-means for tourism product recommendation, which specifically refers to the use of the HHO algorithm to find a set of optimal clustering centers to minimize the distance from the target user to the clustering center. That is, the similarity is maximized.K-means-HHO algorithm decision variable is the clustering center, and Pearson correlation similarity is used as the SO algorithm's fitness evaluation function.

The steps of HHO algorithm combined with K-mean clustering algorithm for travel product recommendation are as follows:

Step 1: Construct a user-tourism product scoring matrix by scoring tourism products with historical user behavior data;

Step 2: Initialize the population. Initialize the number of clustering categories K and randomly initialize the population according to different dimensions to obtain the Harris Hawk population;

Step 3: Calculate the fitness value;

Step 4: Execute the exploration phase and development phase according to the HHO algorithm optimization strategy;

Step 5: Determine whether the K-means-HHO clustering algorithm reaches the maximum number of iterations or satisfies the convergence condition, if yes, output the optimal clustering center; otherwise, loop iterate step 3 to step 5;

Step 6: Output the clustering results and assign the corresponding data to the corresponding categories according to the final clustering results;

Step 7: Pearson correlation similarity is used to calculate the user similarity and identify the set of clusters that have the highest similarity to the target user;

Step 8: Calculate user ratings for uncollected travel products and sort the ratings;

Step 9: Generate recommendation results.

(2) Personalized travel product recommendation model based on XGBoost algorithm

Step 1: Preprocess the text data of reviews of tourism products;

Step 2: Extract the comment text features and construct the dataset;

Step 3: Construct a personalized tourism product recommendation model using the XGBoost method based on the random forest algorithm;

Step 4: Calculate the similarity between the user requirement label and each travel product label and rank the ratings;

Step 5: Generate recommendation results.

(3) Personalized Tourism Product Recommendation Method Flow Combining K-means-HHO Algorithm and XGBoost Algorithm

According to the user-oriented tourism product recommendation based on K-means-HHO algorithm and the item-oriented tourism product recommendation model based on XGBoost algorithm, the fusion produces the recommendation results, and the specific flow chart is shown in Figure 13.

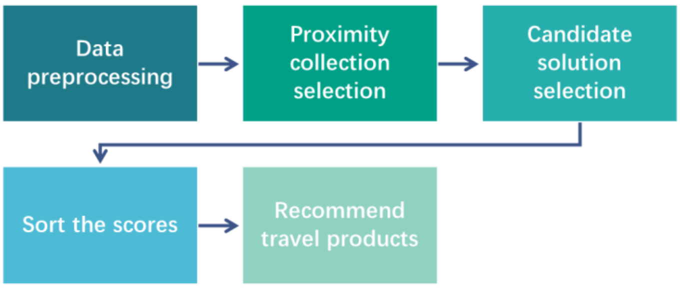

Figure 13 Flowchart of travel E-commerce platform design method based on K-means-HHO and XGBoost algorithm

5 Experiments and analysis of results

5.1 Experimental environment setup

In order to verify the advantages and disadvantages of the travel E-commerce platform design method proposed in this paper, five recommendation algorithms are selected for comparison, and the specific parameters of each algorithm are set as in Table 1.The number of clustering centers of each algorithm's K-means method is determined by Section 5.2. The experimental simulation environment is Windows 10, CPU 2.80GHz, 8GB RAM, programming language Matlab2023a.The data used in this paper comes from the behavioral data of Internet users' travel platform.

Table 1 Parameter settings of tourism product recommendation method

arithmetic	parameterization
K-means	Number of clusters K=4
K-means+RF	N_tree=500, m_try=floor(80.5)
K-means-HHO	The HHO algorithm population is 50 and the number of iterations is 500
K-means-HHO+RF	The HHO algorithm population is 50 and the number of iterations is 500; N_tree=500, m_try=floor(80.5)
K-means-HHO+AdaBoost	The population of the HHO algorithm is 50 and the number of iterations is 500; the number of base classifiers is 10 and the base classifier is RF
K-means-HHO+XGBoost	The population of the HHO algorithm is 50 and the number of iterations is 500; the number of base classifiers is 10 and the base classifier is RF

5.2 Parametric analysis

In order to determine the number of clustering centers of the K-means clustering algorithm, this paper analyzes the analysis of the impact of different number of clustering centers on the performance of the tourism product recommendation method, and the performance analysis of the tourism E-commerce platform design method based on K-means-HHO and XGBoost algorithms under the condition of different number of clustering centers is given in Figure 14 and Figure 15. As can be seen from Figures 14 and 15, this paper analyzes the recommendation accuracy and corresponding time of the recommendation algorithm. As can be seen from Figure 14, with the increase of the number of clustering centers, the personalized tourism product recommendation accuracy gradually increases, and when the number of clustering centers reaches 7, the personalized tourism product recommendation accuracy reaches the maximum. From Figure 15, it can be seen that with the increase in the number of clustering centers, the personalized tourism product recommendation time gradually increases, and when the number reaches 7, the recommendation time tends to be stable. It can be seen that when the number of clustering centers reaches 7, the accuracy of personalized tourism product recommendation is maximum, and the time is no longer increased, therefore, the number of clustering centers is chosen to analyze the subsequent experiments with 7.

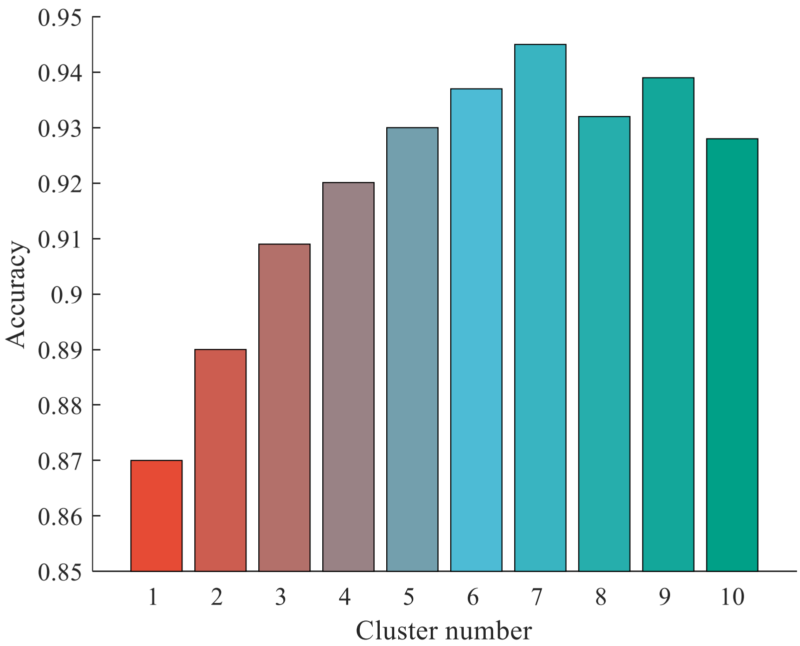

Figure 14 Recommendation accuracy under different number of clustering center conditions

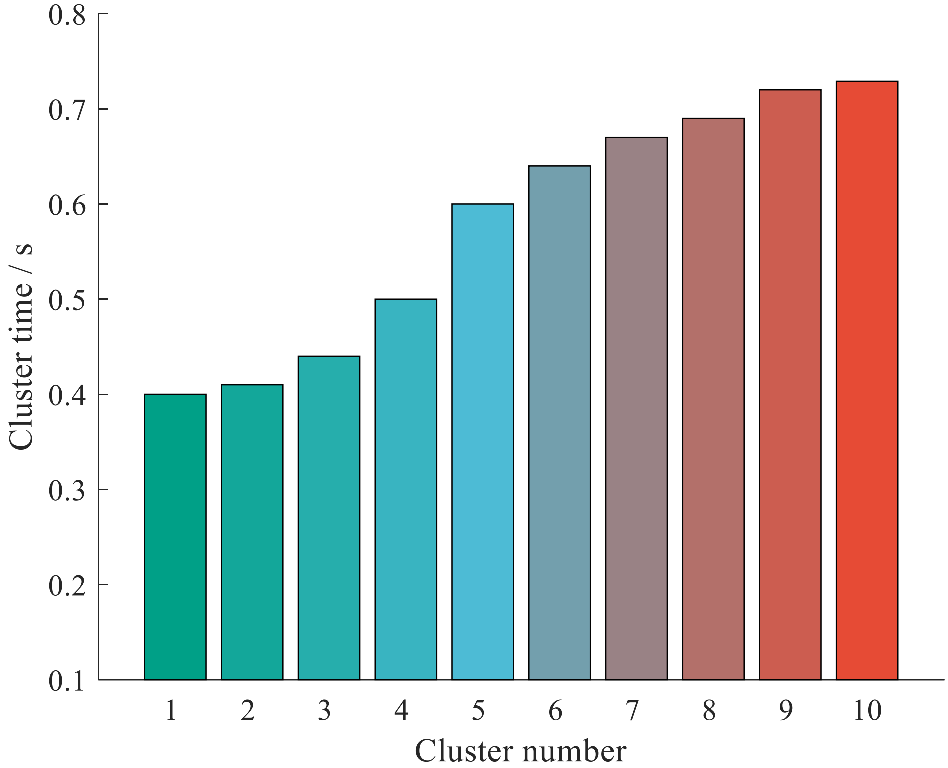

Figure 15 Recommendation response time under different number of clustering centers

5.3 Experimental Performance Analysis

In order to verify the effectiveness and superiority of the travel E-commerce platform design method based on K-means-HHO+XGBoost algorithm, the travel E-commerce platform design method based on K-means-HHO+XGBoost algorithm is compared with K-means, K-means+RF, K-means-HHO, K-means-HHO+RF, K -means-HHO+AdaBoost algorithms for comparison, and the performance results of each model are shown in Figures 16, 17, and 18. 1, 2, 3, 4, 5, and 6 in Figure 16 and Figure 17 represent K-means, K-means+RF, K-means-HHO, K-means-HHO+RF, K-means-HHO+AdaBoost, and K-means-HHO+XGBoost, respectively.

The accuracy of travel E-commerce platform design methods based on each algorithm is shown in Figure 16.From Figure 16, it can be seen that in terms of mean, the accuracy of travel product recommendation method based on K-means-HHO+XGBoost algorithm is the largest, and other algorithms are ranked in the order K-means-HHO+AdaBoost, K-means-HHO, K-means- HHO+RF, K-means+RF, K-means; in terms of standard deviation, the accuracy standard deviation of the travel product recommendation method based on K-means-HHO+XGBoost algorithm is the smallest, and the other algorithms are ranked in the order -means-HHO+AdaBoost, K-means-HHO, K-means-HHO+RF K-means+RF, K-means; comparing K-means-HHO+RF, K-means-HHO+AdaBoost, K-means-HHO+XGBoost algorithms, it can be seen that the personalized travel product recommendation strategy based on XGBoost is better than RF and AdaBoost algorithms; comparing K-means, K -means-HHO, it can be seen that HHO optimization algorithm improves the accuracy of tourism product recommendation method. In summary, the tourism product recommendation method based on K-means-HHO+XGBoost algorithm has the highest accuracy and the best robustness.

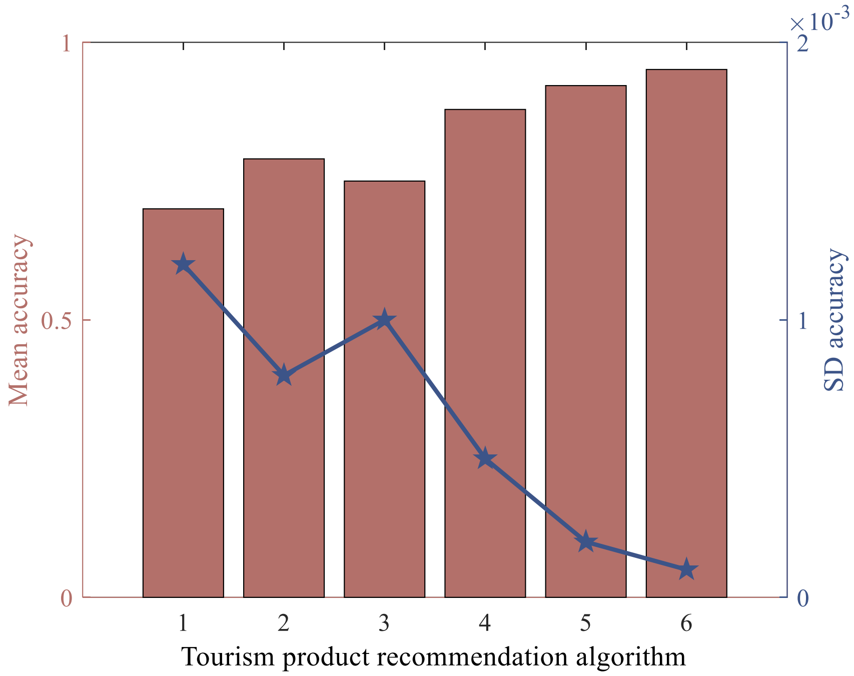

Figure 16 Comparison results of the accuracy of new media event warning methods based on each algorithm

The results of time comparison of travel E-commerce platform design methods based on each algorithm are shown in Figure 17.From Figure 17, it can be seen that in terms of mean value, the travel product recommendation method based on K-means algorithm has the least recommendation time, and the other algorithms ranked in order are K-means+RF, K-means-HHO, K-means-HHO+RF, K-means-HHO+ RF, K-means-HHO+XGBoost, K-means-HHO+AdaBoost; in terms of standard deviation, the recommendation time of the travel product recommendation method based on the K-means-HHO+XGBoost algorithm is the most stable, and the other algorithms ranked in order are K-means, K-means-HHO, K-means- HHO+RF, K-means+RF, K-means-HHO+AdaBoost; by comparing the algorithms, it is found that the HHO optimization algorithm increases the K-means stability. In summary, the recommendation time of the tourism product recommendation method based on K-means-HHO+XGBoost algorithm is 0.68s, which is not the least, but has the best stability and meets the real-time.

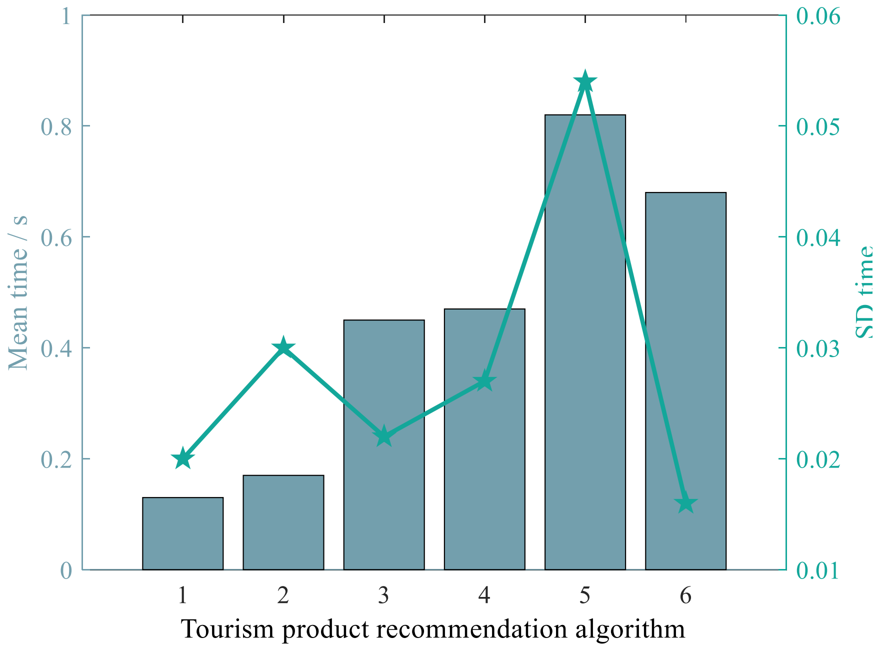

Figure 17 Time comparison of travel product recommendation methods based on each algorithm

The recommendation of travel E-commerce platform design methods based on each algorithm is shown in Figure 18.From Figure 18, it can be seen that, from the aspect of the recommendation rate of browsing behavior, K-means-

HHO+XGBoost results are the largest; from the aspect of the recommendation rate of purchasing behavior, K-means-HHO+XGBoost results are the largest; from the aspect of the conversion rate of recommended traffic to product orders, K- means-HHO+XGBoost has the highest conversion rate.

Algorithms	Browse behavior recommendation rate	Purchase behavior recommendation rate	Referral conversion rate
K-means	0.130	0.207	0.0252
K-means+RF	0.127	0.201	0.0236
K-means-HHO	0.138	0.226	0.0284
K-means-HHO+RF	0.145	0.236	0.0287
K-means-HHO+AdaBoost	0.187	0.289	0.0293
K-means-HHO+XGBoost	0.219	0.341	0.0348

Figure 18 Recommendation of tourism products based on each algorithm

6 Conclusion

With the rise of E-commerce industry and the development of artificial intelligence technology, the way of tourism product recommendation has changed, and tourism E-commerce platform service has become a trend. In order to improve the efficiency of tourism product recommendation and enhance the level of tourism E-commerce services, this paper proposes a tourism E-commerce platform design method based on K-means-HHO and XGBoost algorithm. The method uses the HHO algorithm to optimize the K-means clustering algorithm, constructs a user-oriented tourism product recommendation strategy based on the K-means-HHO algorithm, and at the same time, adopts the XGBoost algorithm to construct an item-oriented tourism product recommendation strategy, and mixes the two strategies to form a tourism E-commerce platform design method. Experimental results show that the recommendation accuracy of the tourism E-commerce platform design method based on the K-means-HHO and XGBoost algorithms reaches more than 0.9, and the recommended response time is 0.67s, of which the accuracy is due to the other algorithms, and it can provide personalized tourism product recommendations for the platform users while satisfying the real-time and enhancing the purchasing volume of tourism products. The HHO in the proposed method in this paper is easy to fall into the local optimum, in the future work, will consider improving the optimization efficiency of the HHO algorithm, making the recommendation effect higher.

Acknowledgements.

This work is supported by Xi'an Traffic Engineering Institute *Research on K-means Algorithm in E-commerce Tourism Platform Design* No. 2024KY-40.

References

[1] Choy M W C , Kamoche K .Factors influencing recommendation of sub-Saharan Africa travel products: a Hong Kong-Kenya importance -performance analysis:[J].Tourism Economics, 2021.

[2] Guixiang Z , Jie C .A Recommendation Engine for Travel Products Based on Topic Sequential Patterns[J]. 2018.

[3] Lin T Y , Chan H T , Hsia C H , Lai C F. Facial Skincare Products' Recommendation with Computer Vision Technologies[J].Electronics, 2022, 11(1):143-.

[4] Zeng D , Liu Y , Yan P , Yang Y. Location-Aware Real-Time Recommender Systems for Brick-and-Mortar Retailers[J].Informs Journal on Computing, 2021(7).

[5] Esmaeili L , Mardani S , Golpayegani S A H , Madar Z Z. A novel tourism recommender system in the context of social commerce[J].Expert Systems with Application, 2020.

[6] Chang J L , Li H , Bi J W .Personalized travel recommendation: a hybrid method with collaborative filtering and social network analysis[J]. issues in tourism, 2022.

[7] Roy D , Dutta M .An Improved Cat Swarm Search-Based Deep Ensemble Learning Model for Group Recommender Systems[J]. Knowledge Management, 2022.

[8] Xing X , Li X .Recommendation of urban vehicle driving routes under traffic congestion: a traffic congestion regulation method considering road network equilibrium[J].Computers and Electrical Engineering, 2023:110.

[9] Wang X .Personalized recommendation framework design for online tourism: know you better than yourself[J].Industrial Management & Data Systems, 2020, 120(11):2067-2079.

[10] Chen L , Cao J , Wang Y , Liang W, Zhu G. Multi-view Graph Attention Network for Travel Recommendation[J].Expert Systems with Applications, 2022, 191. 116234-.

[11] Lee,Hae-Young.What can be expected from the attached travel members? - The moderating role of travel knowledge[J].Korean Journal of Tourism Research, 2011, 26(3):377-393.

[12] Chen L , Wu Z , Cao J , Zhu G, Ge Y. Travel Recommendation via Fusing Multi-Auxiliary Information into Matrix Factorization[J].ACM Transactions on Intelligent Systems and Technology (TIST), 2020.

[13] Nitu P , Coelho J , Madiraju P .Improvising Personalized Travel Recommendation System with Recency Effects[J]. Big Data Mining and Analytics(English), 2021(3):16.

[14] Chen L , Cao J , Liang J Y Q .Keywords-enhanced Deep Reinforcement Learning Model for Travel Recommendation[J].ACM transactions on the web, 2023, 17(1):1.1-1.21.

[15] Zhu Z , Wang S , Wang F , Tu Z. Recommendation networks of homogeneous products on an E-commerce platform: measurement and competition effects[J]. Expert Systems with Application, 2022(09):201.

[16] Manwade P P N K B .Survey on Recommendation of E-commerce products to existing and new user by Analyzing Social data along with E-commerce data[J]. International Journal for Modern Trends in Science and Technology, 2020, 6(7):78-84.
[17] Tang Q , Zhong S .A Personalized Travel Route Recommendation Model Using Deep Learning in Scenic Spots Intelligent Service Robots[J]. robotics, 2022(Pt.1):2022.
[18] Chang L V .Recommendation and Sentiment Analysis Based on Consumer Review and Rating[J]. , 2020, 11(2).
[19] Wan H , Li Y .A personalized recommendation algorithm of user preference products based on Bayesian network[J]. Development, 2021, 25(2):85-.
[20] Yoon S , Lee T H .The Effect of the Perceived Lovemark Constructs of Travel Agency on Repurchase/Recommendation Intention and Consumer Support toward the Extension of Discounted Travel Products[J].Journal of Korea Service Management Society, 2013, 14(1):49-72.
[21] Cui C , Wei M , Che L , Wu S, Wang E. Hotel recommendation algorithms based on online reviews and probabilistic linguistic term sets[J].Expert Systems with Application, 2022.
[22] Xu B , Chen Z , Wang X , Bu J, Zhu Z, Zhang H. Combined prediction model of concrete arch dam displacement based on cluster analysis considering signal residual correction[J].Mechanical Systems & Signal Processing, 2023:203.
[23] Chen G , Cockburn B , Singler J R , Zhang Y. Superconvergent Interpolatory HDG Methods for Reaction Difusion Equations II: HHO-Inspired Methods[J]. Journal of Applied and Computational Mathematics (in English), 2022, 4(2):23.
[24] Ji W , Liao Y , Zhang L .Integrated Generative Adversarial Network and XGBoost for Anomaly Processing of Massive Data Flow in Dispatch Automation Systems[J]. Intelligent Automation and Soft Computing(English), 2023, 37(9):2825-2848.
[25] Shen Weilei, Zhang Xinyang, Yang Xuechun. Research on quality control method of clutch manufacturing process based on random forest[J]. Journal of Hefei University of Technology (Natural Science Edition), 2023(11):1441-1446,1500.