Research on User Profile Combined with Collaborative Filtering Recommendation Algorithm for Intelligent Tourism

Chun He, Caijian Hua

School of Computer Science and Engineering, Sichuan University of Science & Engineering, Sichuan YiBin 643000, China

Abstract: In recent years, the online travel sector in the tourism industry has experienced significant growth and popularity due to the development and widespread adoption of internet technology and smart devices. However, despite these advancements, scenic spots have struggled to provide precise services to tourists, as the online marketplace is flooded with numerous and disorganized commodity resources, lacking standardized construction and systematic management. As a result, travelers find it challenging to access specialty goods that cater to their personalized needs. To address this issue, this paper proposes the use of user profiling and collaborative filtering recommendation algorithms to achieve personalized recommendations for specialty products in scenic spots. The general process of constructing user profiles for scenic spots and combining them with collaborative filtering algorithms to create an intelligent tourism recommendation system is outlined. The paper also highlights the current challenges faced by this system in practical applications and provides future research prospects to promote accurate services in tourist attractions.

Keywords: User profiling, Recommendation, Personalization, Tourism.

1. Introduction

In recent years, with the rapid development of the tourism industry, personalized travel service models and demands have been continuously strengthened, becoming a new focal point in the tourism market. Consumers are placing increasing importance on the quality of their travel experiences, seeking uniqueness and personalized adventures, and are willing to pay higher costs for them. Moreover, as the interest in travel grows among young people, they prefer independent and personalized travel styles, making personalized travel services and related products and offerings more popular. Personalized travel service is a consumer-oriented model, aiming to meet individualized needs by providing comprehensive, high-quality, and personalized tourism services through refined management and differentiated services.

Consumers can choose travel routes, itineraries, and activities that suit their preferences while receiving professional advice and tailored services. This approach meets consumers' more personalized demands and offers a wider array of choices. However, the current scenic spots have not effectively addressed the personalized needs of travelers. Disorganized goods still flood major travel platforms, leaving tourists in a dilemma when selecting souvenirs and specialty products from scenic spots. To address this, in order to meet the increasingly urgent and efficient demand for intelligent travel information and effectively tackle the problem of information overload in the tourism industry, online travel service platforms urgently need to introduce intelligent and personalized travel recommendation services through continuous technological innovation. These services can provide users with more diverse and effective travel information and personalized recommendations for attractions, assisting users in decision-making and itinerary planning.

On April 6, 2023, the Ministry of Culture and Tourism of China issued the "Notice on Strengthening the Collaborative Innovation and Development of 5G+ Smart Tourism." This initiative was launched to deeply implement the "14th Five-Year Plan for National Economic and Social Development and the Vision 2035," "14th Five-Year Plan for the Development of the Information and Communication Industry," "14th Five-Year Plan for the Development of the Tourism Industry," "5G Application 'Sailing' Action Plan (2021-2023)," and the "Opinions on Deepening 'Internet + Tourism' to Promote High-Quality Development of the Tourism Industry." The goal of this notice is to promote the innovative application of 5G technology in the tourism industry, encouraging market entities to explore 5G integration and innovation applications. The focus is on enhancing tourism services, improving the travel experience, innovating tourism management, optimizing tourism resources, and conducting business and model innovations. In line with this initiative, efforts will be made to nurture a group of 5G+ smart tourism solution providers and innovative enterprises. This will empower the scale development of 5G+ smart tourism innovative applications. Ultimately, this policy paves the way for personalized smart tourism, aiming to meet the individualized needs of travelers with advanced 5G-powered solutions.

User profiling refers to the process of collecting, organizing, and analyzing various data about users to construct comprehensive characteristics and descriptions, unveiling their interests, preferences, behavior patterns, and needs[1]. It serves as a fundamental cornerstone for personalized services and decision-making in recommendation systems, targeted advertising, and precision marketing. The primary objective is to assist businesses in gaining a better understanding of their users and providing more personalized and accurate products and services. On the other hand, collaborative filtering algorithms are a commonly used category of algorithms in recommendation systems[2]. They analyze user behavior and preferences to discover
similarities between users or items, thereby generating personalized recommendations for users. Built on the concept of "collective intelligence," collaborative filtering algorithms infer user behavior and preferences based on the shared behaviors of groups. Through the labels generated from user profiling, these algorithms can achieve personalized and precise product recommendations for tourist attractions[3].

In conclusion, this article aims to establish personalized tourism services with strong support from national policies. It lays a solid foundation for the personalized tourism ecosystem by employing user profiling techniques to construct user models and combining them with collaborative filtering algorithms to create a recommendation model for scenic spot specialties.

2. Related Work

In this chapter we discuss what is user portrait and the research status of smart tourism and collaborative filtering algorithm.

2.1. User portrait overview

Early user portraits were based on traditional market research and consumer behavior analysis. Marketers collect basic information and purchase behavior of users through telephone surveys, mailing questionnaires, etc. to understand user needs and preferences, and then use this information for advertising and product positioning[4].

With the development of the Internet and digital technology, user portraits have entered a new stage. The popularity of the Internet has made it possible to obtain a large amount of user behavior data, such as users' browsing records on websites, click behaviors, search habits, etc. These data provide more dimensional features for user portraits[5]. The rise of machine learning and big data technology has made the modeling and analysis of user portraits more efficient and accurate. By using various machine learning algorithms and models, users' interests and behavior patterns can be automatically learned from massive data, so as to construct user portraits more accurately. In recent years, the development of deep learning technology has further promoted the evolution of user portraits. Deep learning can process more complex user behavior data and text content, and extract deeper user characteristics, so as to achieve higher levels of personalized recommendation and precision marketing[6].

2.2. Research Status of Collaborative Filtering Algorithm

In 1992, Paul Resnick and others proposed an early collaborative filtering algorithm and applied it to the recommendation of Usenet newsgroups. Based on the similarity between users, this method recommends items that are liked by similar users, thereby realizing personalized recommendation[7]. Subsequently, the collaborative filtering algorithm has received extensive attention and research, which has triggered an upsurge in the field of recommendation systems. Researchers have proposed various improvements and extensions, including item-based collaborative filtering, latent semantic models, matrix factorization, etc., to improve the accuracy and efficiency of recommendations. In the recommendation algorithm competition held by Netflix in 2006, the famous Matrix Factorization algorithm (Matrix Factorization) made a significant breakthrough in the collaborative filtering algorithm, making the collaborative filtering algorithm more mature and practical.

In the application of collaborative filtering algorithm, Gao Chaomeng and Wang Yonggang[8] used collaborative filtering algorithm to analyze and study brand visual design in order to improve their own brand building level, and proposed a collaborative filtering algorithm to analyze the visual design of corporate brands due to the low accuracy of general recommendation. Communicate the design process. Quickly find the books you want, and propose a book recommendation algorithm based on collaborative filtering. Wang, Zhi Hui and Hou, De Zhij[9] take the interestingness of the book itself as an important measurement index, including search times, borrowing time, borrowing times, borrowing intervals, renewal times, etc. to improve the search rate. Wang, Tianyu and Ge, Dong[10] built a recommendation system based on a collaborative filtering multi-dimensional hybrid algorithm to find, analyze and recommend online Chinese learning resources that meet the needs of learners and teachers.

2.3. Application of Collaborative Filtering Algorithm in Smart Tourism

In terms of collaborative filtering applied to smart tourism, in order to find one’s favorite spots in a large number of tourist attractions, Niu, Tianjiao, Song, Mei, etc.[11] have improved the collaborative filtering algorithm based on the low accuracy of recommendation considering changes in user interests. The time function of Ebbinghaus' law of forgetting is used to calculate the user similarity to ensure the accuracy of the recommendation. Wang and Zhonghua[12] proposed that the intelligent recommendation model of tourist destinations requires an algorithm that can accurately recommend tourist attractions according to the interests of users. However, traditional collaborative filtering algorithms ignore the problem of user preferences, which may lead to a decrease in recommendation accuracy. To address these issues, he analyzed the factors that affect users' interest preferences based on their global and local rating information. By calculating the global probability distribution of user rating information, the Hamming approximation is used to calculate the user's interest preference. The similarity algorithm is effectively combined with the traditional similarity algorithm to propose a collaborative filtering recommendation algorithm model based on user preferences for tourist attractions under sparse data.

3. Methodology

In this chapter, we discuss the technical flow and design of user profile and collaborative filtering algorithm.

3.1. The overall architecture of the model

This paper proposes to build a model of user portraits to generate passenger tags, associate labels with items, extract corresponding tags or feature tags for the features of each scenic spot specialty, and calculate the similarity between user profile tags and scenic area specialty products. Select the group of users that are most similar to the target users to generate recommended columns for them. The conceptual model is shown in figure 1.
3.2. Construction of user portrait model

The virtual user portrait is based on real data, and the result is a differentiated label description of different users after being segmented. User tags have semantic and short text features. The following mainly introduces the process of obtaining tags from passenger text data. Process the obtained scenic spot data through source data collection and data preprocessing, and complete the entire user portrait through the process of behavior modeling and user portrait creation.

① Source data collection. Static data often comes from the information filled in by users on the website. When there is no data, it can be predicted by building a model. For example, when the user's gender information is unknown, a model can be built to judge based on the user's purchase behavior. Dynamic data refers to the data generated by user behavior, such as browsing scenic spots pages, purchasing scenic spot products, collections, evaluations, etc. These behavioral data can calculate the customer's brand preference, consumption ability, order quantity, category purchase ranking and other information.

② Data preprocessing. When the data is collected, it is generally necessary to do some preprocessing on the data, such as data cleaning and data structuring. While removing some dirty data, the structure of the data can also be standardized for better subsequent processing and analysis.

③ Behavioral modeling. The user's static data and dynamic data are calculated through different dimensions of statistics and algorithm models, and the data is labeled to express the user's interests, needs, preferences, etc.

④ User portrait creation. Deepen the behavior modeling process, and use prediction and Kmeans clustering algorithms to complete the mapping of different attributes of users to tags. The completion of user labeling means that the user's portrait is basically completed.

Then build a user recommendation subsystem based on the data model of user portrait analysis to realize personalized recommendations for user products and scenic spots.

The user portrait model architecture diagram of its smart tourism platform is shown in Figure 2:
### 3.2.1. Chinese participle

In the process of building user portraits, there are spaces between each English word as a very obvious separator\[^{[13]}\], while Chinese words are continuous, which cannot be understood and processed by the computer. In order to make the Chinese text data generated by passengers into meaningful Words or phrases, so as to better understand and analyze user behavior and interests. This article will perform Chinese word segmentation processing on the obtained data, so that a piece of text can be divided into independent words. Commonly used Chinese word segmentation methods include dictionary-based methods, statistics-based methods, and deep learning-based methods.

This article uses Jieba word segmentation as a tool for text processing. It combines word segmentation methods based on string matching and statistics. The word segmentation effect is better. The word segmentation process is as follows:

1. Initialize the tokenizer.
2. According to the characteristics of tourists, the scenic spot can customize the dictionary to contain specific words or terms.
3. Use the forward maximum matching algorithm for word segmentation. The algorithm starts at the beginning of the text and matches the longest word. If a word is matched, the word will be used as a word segmentation result; if no word is matched, the current character will be output as a word.
4. After the forward maximum matching word segmentation, process some rules and normalization, such as merging first and last names, removing single words, etc.
5. The word segmentation is completed, and the stuttering word segmentation is used to output the results.

#### 3.2.2. Feature word extraction

The main purpose of extracting feature words from user portraits is to convert passengers’ complex behaviors and interests into limited and meaningful features in order to better describe and understand users’ characteristics and behavior patterns\[^{[14]}\]. Through feature word extraction, these behavioral data can be converted into a small number of meaningful feature words, thereby reducing the dimension and complexity of the data; it can accurately reflect the concerns of passengers, ignoring irrelevant or noise information; the characteristics of passengers and Behaviors are transformed into feature vectors, thereby realizing personalized recommendation of special products in scenic spots and calculation of passenger similarity. This paper uses TF-IDF as the main method of feature word extraction, counts the frequency of each word in the text, and multiplies the frequency of a word with the frequency of documents appearing in the text collection to get a more representative word. The value of importance in the text. The calculation process of TF-IDF is as follows:

1. Calculate word frequency (TF)
   \[ TF = \text{Occurrences of special words} \]
2. Calculate inverse document frequency (IDF)
   \[ IDF = \log \frac{SUM}{N + 1} \]

Where SUM represents the total number of documents in the corpus, and N represents the number of documents containing the word.
3. Calculate TF-IDF
   \[ TF_{IDF} = TF \times IDF \]

#### 3.2.3. Label design

The design of the user portrait label system can fully restore the user’s characteristics, and improve the efficiency of computer analysis and processing of labels\[^{[15]}\]. It is mainly divided into basic attribute tags and behavior feature tags. The basic attribute tags describe the basic information and characteristics of the user, and the behavior feature tags describe the user's behavior habits and behavior patterns. The user portrait label system of the smart tourism platform constructed in this paper is shown in Figure 3.

Among them, the label represents the user's interest point, and the label weight is designed to represent the user's preference. According to TF-IDF, the user's weight value for the tag can be obtained, but the weight at this time only considers the relationship between the user and the tag, and the user's dynamic tag is constantly changing, so consider the business scene where the tag is located, the occurrence elements such as the time interval of the tag, the number of times the user generates the tag, etc. Therefore, this paper sets the user's dynamic label weight as the product of the behavior type weight (P), time decay (T), behavior times (N) and the weight (TF) obtained by TF-IDF. P focuses on the degree of collection of special products in scenic spots, T indicates the affected behavior, TF is the value obtained by TF-IDF, and the greater the data of N, the greater the corresponding weight. At the end of time, according to the weight ranking of the obtained keywords, the nouns with the highest weight value are extracted as the user's feature labels, and the similarity between users can be obtained by calculating the label similarity between users (S).

\[ S = P \times T \times TF \times N \]

#### 3.2.4. Cluster analysis

The purpose of cluster analysis in the process of user portrait modeling is to group similar passengers into one category, thereby forming passenger groups or passenger clusters, and each group represents passengers with similar characteristics and behaviors. This facilitates the classification and segmentation of different travelers to better

---

**Figure 3.** Smart tourism platform user portrait label system
understand the needs and behavior patterns of different groups of travelers. Common cluster analysis methods include K-Means Clustering, Density-Based Clustering, Spectral Clustering, etc.

The text uses the K-Means Clustering method to cluster the passenger data, and iteratively finds K clusters to minimize the loss function corresponding to the clustering results. Among them, the loss function is defined as the sum of the squares of errors between each passenger sample and the center point of the cluster to which it belongs:

\[ J(c, u) = \sum_{i=1}^{M} ||x_i - u_c||^2 \]

Where \( x_i \) represents the i-th passenger sample, \( c_i \) is the cluster to which \( x_i \) belongs, \( u_c \) represents the center point corresponding to the cluster, and \( M \) is the total number of samples. The passenger clustering process is as follows:

1. Data preprocessing: standardization, outlier filtering.
2. Randomly select K centers, denoted as \( u_1^{(0)}, u_2^{(0)}, \ldots, u_K^{(0)} \).
3. Define the loss function:

\[ J(c, u) = \sum_{i=1}^{M} ||x_i - u_c||^2 \]

(4) Let \( t=0,1,2,\ldots \) be the number of iteration steps, repeat the following process until J converges:

4.1 For each sample \( x_i \), assign it to the nearest center \( c_1^1 < -\text{argmin}_u ||x_i - u_1||^2 \).

4.2 For each class center \( k \), recalculate the center of the class

\[ u_k^{(t+1)} = -\text{argmin}_u \sum_{i: c_i^{(t)} = k} ||x_i - u||^2 \]

Finally, it is divided into K types of groups through cluster analysis. For each cluster, the commonality and characteristics of users can be analyzed, and one or more labels can be assigned to the cluster according to the designed label system. These tags can reflect the interests, behavior habits and other characteristics of the user group. Finally, the labels assigned to each cluster are associated with the original user data to construct a user group model.

3.3. User portrait model combined with collaborative filtering algorithm

The collaborative filtering algorithm can realize the personalization and accuracy of product recommendation through the tags generated by user portraits\[16\]. The tags in user portraits are descriptions of passenger interests, preferences, and behaviors, which can help collaborative filtering algorithms better understand users and make personalized recommendations. In this paper, each user group constructed by the user portrait is used as a data source, and a user-based collaborative filtering algorithm is used for recommendation. In the collaborative filtering algorithm, for the target passenger, find other users who are similar to them in the user portrait model, and according to the behavior and preferences of these similar users, recommend products that may be of interest to the target user and complete the personalized travel recommendation for passengers. The following is how the collaborative filtering algorithm uses the user group generated by the user portrait to realize the product recommendation process:

1. User group feature value extraction

For each user group, group features are further extracted. In this paper, user age and passengers' different ratings on commodities are used as group features. These group features can help to better understand the commonality and needs of user groups.

2. Collaborative filtering algorithm

When implementing personalized recommendation, user-based collaborative filtering algorithm is used. For each user group, using the similarity calculation method, this paper uses the Pearson correlation coefficient to calculate the similarity, and find other user groups that are most similar to this group, that is, similar groups. The algorithm formula of the Pearson correlation coefficient is as follows:

\[ r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}} \]

Among them, \( x_i \) takes the age of passengers as the behavior variable, \( \bar{x} \) is the average age of passengers, \( y_i \) takes the ratings of passengers on different commodities as data variables, \( \bar{y} \) is the average rating of commodities, and \( n \) is the sample size of passenger group data.

3. Group recommendation

Using the behavior and interests of similar groups as a reference, recommend items that may be of interest to the target user group. The recommendation results of similar groups can be weighted and averaged, or the recommendation results can be combined based on diversity.

4. Generate recommendation list

According to the list of items recommended by the group, the final recommendation list is generated with the predicted score of the item as the weight, and presented to the target user group.

The flow chart of its user portrait combined with collaborative filtering algorithm is shown in Figure 4:

![Figure 4. Flow chart of user portrait combined with collaborative filtering algorithm](image-url)

4. Result Analysis

In this chapter, we compare the test standards and test results of the traditional collaborative filtering algorithm and the user portrait combined with the collaborative filtering
The collaborative filtering recommendation algorithm combined with user portraits is aimed at a large number of mobile users in scenic spots, and solves the problem that it is difficult to make accurate predictions and recommendations due to the lack of behavior data of new tourists. In this paper, new users are divided into user groups, and the recommendation list constructed based on the behavior characteristics and attribute characteristics of old users is used as the basis to achieve the purpose of accurate and personalized recommendation of special products in scenic spots for new and old users of scenic spots. In order to ensure the authenticity and scientificity of the data, this paper collects real data based on tourism websites as the old user data set, and manually adds new user passenger information as the new user data set.

### 4.1. Measured standard

There are three evaluation indicators in the smart travel recommendation system, which are accuracy, novelty and coverage [17]. Accuracy is a test of the results of an algorithm. Accuracy is used as the index to evaluate the experiment. The experiment uses MAE as the evaluation standard for test verification. MAE reflects the gap between the actual rating value and the predicted rating value [18]. The smaller the gap between the two, the better the recommendation effect. It is assumed that the predicted rating set of the user portrait combined with the collaborative filtering model is \( \{p_1, p_2, ..., p_n\} \), the user group's rating set for the product is \( \{q_1, q_2, ..., q_n\} \), the calculation formula of MAE can be expressed as:

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |p_i - q_i|
\]

Among them, \( N \) is the size of the test set, \( p_i \) is the predicted rating value given by the recommendation algorithm, and \( q_i \) is the actual average rating of the item commodity by the actual passenger group. The larger the error of MAE, the larger the value. The smaller the value of MAE, the better the accuracy of the prediction model.

### 4.2. Result analysis

The results of the test set are shown in Figure 5. Under the condition of dividing the scale of passenger set the number of groups = 10, 20,...,60, it compares the accuracy of the traditional collaborative filtering algorithm when testing the same number of new passengers.

![Figure 5. Comparison of algorithm test set results](image)

It can be seen from the change trend of MAE in the figure that the MAE of the two algorithms will change with the number of neighbors of the passenger group. When the number of neighbors is small, because only a few people's opinions are referred to, the prediction score is greatly affected by personal factors. Compared with passengers who only have attribute characteristics, the reference value of neighbors is small, resulting in their MAE larger, the prediction accuracy is lacking. The collaborative filtering algorithm combined with user portraits is grouped under similar attribute features, based on the behavior characteristics of old users, and by referring to the data of the neighbors of the user group, the more the number of neighbors of the passenger group within a certain range, the smaller the MAE, and the better the prediction results. Accurate. From the above analysis, the collaborative filtering algorithm combined with user portraits proposed in this paper is superior to traditional collaborative filtering algorithms in terms of recommendation performance for new passengers and users without behavioral characteristics.

### 5. Conclusion and Discussion

Travel recommendation is a specific application field of intelligent recommendation system, and collaborative filtering recommendation technology is the most commonly used recommendation algorithm. This paper proposes to generate user tags by building a model of user portraits, and associate tags with items. For the feature extraction of each scenic spot specialty, I want to extract the corresponding tags or feature tags, and calculate the similarity between the tags in the user portrait and the scenic spot specialty tags, to measure the user's interest in the item, based on the similarity between the user and the special products of the scenic spot, select a group of users who are most similar to the target user to generate a recommendation list for it, so as to solve the problem of excessive user base in collaborative filtering recommendation sparsity problem.

In this way, the collaborative filtering algorithm uses the tags generated by user portraits to achieve personalized and accurate product recommendations. It can better understand the interests and needs of passengers, so as to provide passengers with product recommendations that meet their personalized preferences, and improve passenger satisfaction and purchase experience. At the same time, based on the findings of this paper, there are still some directions for improvement in the future:

1. **Solve the cold start problem**: When a new user joins the system or a new item is launched, due to the lack of user behavior data or item rating information, traditional collaborative filtering algorithms and user portraits may not be able to make accurate recommendations. Solving the cold start problem requires introducing other information sources, such as content information, social network data, etc., to help establish portraits and recommendation models for new users or new items.

2. **You can add context information**: The current user portrait combined with collaborative filtering algorithm is mainly based on the user's historical behavior and characteristics, but ignores the contextual information when recommending, such as time, geographical location, device, etc. Integrating contextual information into the recommendation model can better adapt to the current needs and situations of users and provide more accurate recommendations.
3. By integrating multiple recommendation algorithms: Although the collaborative filtering algorithm performs well in personalized recommendation, it is not suitable for all scenarios. Collaborative filtering algorithm can be integrated with other recommendation algorithms (such as content recommendation, deep learning model, etc.) to form a hybrid recommendation system to make full use of the advantages of different algorithms and improve the diversity and accuracy of recommendations.

4. User privacy protection: User portraits involve the user's personal information and behavioral data, and it is necessary to protect the user's privacy and data security. In the recommendation model, privacy protection measures should be strengthened, such as adding a blockchain network architecture to ensure data security and prevent user portraits from being abused or leaked.

References


