



Applied Artificial Intelligence

An International Journal

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/uaai20>

Intelligent recommendation model of tourist places based on collaborative filtering and user preferences

Zhonghua Wang

To cite this article: Zhonghua Wang (2023) Intelligent recommendation model of tourist places based on collaborative filtering and user preferences, Applied Artificial Intelligence, 37:1, 2203574, DOI: [10.1080/08839514.2023.2203574](https://doi.org/10.1080/08839514.2023.2203574)

To link to this article: <https://doi.org/10.1080/08839514.2023.2203574>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 21 Apr 2023.



Submit your article to this journal [↗](#)



Article views: 377



View related articles [↗](#)



View Crossmark data [↗](#)



Intelligent recommendation model of tourist places based on collaborative filtering and user preferences

Zhonghua Wang

School of Management, Xuzhou University of Technology, Xuzhou, Jiangsu, China

ABSTRACT

Providing personalized recommendation service for users and improving the accuracy of recommendation and user satisfaction are the main research tasks of current travel recommendation systems. The intelligent recommendation model of tourist places requires the algorithm to be able to accurately recommend tourist attractions according to the user's interests. Big data and artificial intelligence technologies have driven the development of intelligent recommendation systems. However, realistic data are often sparse, and the lack of common rating items among users makes some traditional similarity measures impossible to calculate. In addition, traditional collaborative filtering algorithms ignore the issue of user preferences, which can cause a decrease in recommendation accuracy. To address these issues, this paper analyzes the factors affecting users' interest preferences in terms of their global and local rating information. The interest preferences of users are calculated by computing the probability distribution of user rating information globally and using Hamming approach degree. A similarity algorithm about user preferences is derived using Jeffries-Matusita distance. The similarity algorithm is effectively combined with the traditional similarity algorithm to propose a model of collaborative filtering recommendation algorithm for tourist attractions based on user preferences under sparse data. The paper aims to improve the accuracy of tourist attraction recommendation systems by incorporating user preferences and addressing the issue of sparse data and the lack of common rating items among users that traditional similarity measures cannot calculate. The experimental results show that the improved algorithm model outperforms the traditional collaborative filtering algorithm and other algorithms. And it also has high accuracy rate on more sparse tourism data set.

ARTICLE HISTORY

Received 15 March 2023

Revised 11 April 2023

Accepted 12 April 2023

Introduction

With the development of the domestic economic level and the improvement of living standards, China's tourism industry pattern continues to improve, the market scale is also expanding, China has gradually entered the era of mass

CONTACT Zhonghua Wang wzh_js@126.com School of Management, Xuzhou University of Technology, Xuzhou, Jiangsu 221018, China

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

tourism. In 2019, China's tourism market demand continues to improve, with stable growth in domestic tourism, stable development of the outbound tourism market, and an increase of 11.4070 in the number of domestic tourists over the same period of the previous year. tourism-related industries are gradually becoming the pillar industry of China. The booming development of the tourism industry has also brought about changes in the form of consumption and consumption demand: from the traditional "tour with the group" to the hot "self-driving tour," "peripheral tour" and "rural tour." The changes in the form of tourism consumption, such as "rural tourism," reflect that tourists have higher requirements for the infrastructure and service level of tourism areas, and higher requirements for personalized tourism products and services, and tourists pay more attention to improving tourism quality and user experience through personalized consumption. So, providing tourists with personalized tourism services is the future development trend of tourism (Hao and Song 2021).

In recent years, the rapid development of the Internet and tourism has broken the traditional tourism situation. A variety of travel online service e-commerce platforms are rapidly emerging, providing more information and choices for tourism travel. Travel products are gradually shifting from traditional offline to online markets, and the proportion of people choosing travel locations and making travel plans with the help of online travel platforms will further expand in the future. As more and more people use online travel platforms and share travel information on them, online travel information is growing explosively (Cui, Luo, and Wang 2018). Faced with the huge and complex travel data people are easily overwhelmed by the large amount of information. And the basic search function provided by the current online travel service platform is difficult to help users quickly find the travel information they need and choose their preferred travel destinations. Therefore, in order to meet the increasingly urgent demand for efficient and intelligent travel information and effectively solve the problem of travel information overload, online travel service platforms urgently need to introduce intelligent and personalized travel recommendation services through continuous technological innovation to provide users with more and more effective travel information and personalized recommended attractions to help users make decisions and develop itineraries.

Personalized travel place recommendation aims to use personalized recommendation technology to tap into users' travel preferences and potential deep needs and provide travel recommendations for users. Not only can it help users discover interesting and fun travel locations, make travel plans, etc., but also help people discover friends with similar travel interests. In addition, personalized travel recommendation is also of great relevance to online travel platforms, which can help online travel platforms improve user satisfaction and economic efficiency. Personalized recommendation

technology, as a research hotspot, has been widely studied and applied in various other fields. For example, in the field of e-commerce Amazon uses recommendation system to recommend products to users to increase sales (Smith and Linden 2017). Netflix uses personalized recommendation to recommend movies to users (Floegel 2020). Most of the videos, news and advertisements on software such as Shake and Today's Headlines come from recommendations. In the academic field personalized recommendations can help us find the needed literature quickly (N, O, and O 2022). In the social field personalized recommendation can help us discover interesting friends (Zhang et al. 2020).

Recommendation systems have been widely used to recommend books, articles, movies, TV shows, news, music pages, etc. The earliest concept of recommendation system is to use e-commerce websites to provide customers with product information and suggestions to help users decide what products they should buy, simulating a salesperson helping customers through the buying process. 1997 AT&T Labs proposed a personalized recommendation system based on collaborative filtering algorithm, which is widely used in large e-commerce websites such as eBay, YouTube, etc. Many scholars have researched and designed travel recommendation systems based on data mining techniques, LBSN (location-based social network) data, context-aware techniques, constraints, vertical search engines, content-based recommendation techniques and hybrid recommendation techniques, weighted association rules, ontologies, 3 G cell phones, cloud computing, and the Internet of Things to recommend travel packages, travel products, travel routes, tourist attractions, and itineraries that meet the needs of travelers (Memon et al. 2015).

The development of mobile Internet generates huge amount of travel information, and many travel service providers have launched travel recommendation systems to help tourists solve the problem of information overload (Du et al. 2019; Lin et al. 2020). Collaborative filtering algorithm is a widely used technique in travel recommendation systems nowadays, which analyzes users' needs and user history record data to obtain users' preference characteristics, predicts interested travel information for users, and then recommends the information to users. However, in the traditional travel recommendation system, it is mostly based on the user's geographic location information and the TOPN of tourist attractions to generate recommendation information for users. This can lead to tourists gathering to the same tourist attractions because of similar travel information, which can lead to tourism problems such as overflowing tourists in Mount Tai and fire in Mount Wutai in Shanxi. And the traditional travel recommendation system ignores the tourist's consumption level and travel time requirements. For example, the travel recommendation system recommends high consumption travel routes for student groups or generates

recommendations that lead to conflicts between tourists' working hours and travel time, which leads to the final recommendation results not meeting tourists' individual needs.

Realistic data is usually sparse. Under the sparse data, the similarity calculation will become inaccurate. To address this problem, this paper proposes a collaborative filtering research algorithm based on user preferences under sparse data from a global perspective, digging deeper to analyze the factors affecting user interest preference attributes and proposing a collaborative filtering research algorithm under sparse data. The improved algorithm considers the following points: 1) Different users have different interest preferences. Even in more sparse data, this preference feature can be reflected. The accuracy of the recommendation can be improved to a great extent by grasping the users' preferences. 2) Each user has his own rating style. Some users may not like to give high scores, and some users may give the same scores to different contents. This paper considers the impact of user rating preferences on recommendations. 3) Improving the algorithm requires considering the ratio of absolute user ratings to common ratings.

Specifically, the paper aims to improve the accuracy of tourist attraction recommendation systems by incorporating user preferences and addressing the issue of sparse data and the lack of common rating items among users that traditional similarity measures cannot calculate. The research analyzes the factors affecting user interest preferences using global and local rating information. The paper proposes a collaborative filtering recommendation algorithm for tourist attractions based on user preferences, which uses a similarity algorithm based on Jeffries-Matusita distance to address the aforementioned challenges. It proposes an improved collaborative filtering algorithm for tourist attraction recommendations, which effectively combines the traditional similarity algorithm with the similarity algorithm based on user preferences. The proposed algorithm outperforms traditional collaborative filtering algorithms and other algorithms, achieving high accuracy even on sparse tourism data sets. Finally, the experimental results show that the proposed algorithm model outperforms traditional collaborative filtering algorithms and other algorithms. The use of user preferences and Jeffries-Matusita distance proves to be effective in improving the accuracy of recommendation systems in the tourism domain, making it a promising approach for future research in this field.

The rest of the article may be structured as follows:

- (1) Related Works: This section presents a review of previous research on collaborative filtering algorithms for recommendation systems and user preference modeling, highlighting the gaps and limitations of existing approaches.

- (2) Collaborative Filtering Algorithm Based on User Preference: This section describes the proposed algorithm in detail, including the methods for calculating user interest preferences, the similarity algorithm based on Jeffries-Matusita distance, and the approach for combining this algorithm with traditional similarity measures.
- (3) Experiments: This section presents the experimental design and results, evaluating the performance of the proposed algorithm on various datasets and comparing it with other collaborative filtering algorithms.
- (4) Discussion and Conclusion: This section summarizes the key findings and contributions of the research, discusses the implications of the results, and suggests directions for future research in this area. It also highlights the limitations of the proposed approach and areas for further improvement.

Related Works

Research Status of Tourism Recommendation System

With the rapid development of artificial intelligence (Chen et al. 2022; Chen, Zeng, and Li 2022) and information technology (Chen, Jiang, and Fu 2021), the demand for deeper integration between the tourism industry and information technology has become stronger and stronger. In the era of big data with information overload, users often feel that they cannot get the information they are interested in quickly. Therefore, personalized travel recommendation technology is born to recommend users' interests according to their preferences (Bai, Pamula, and Kumar Jain 2019). Personalized travel recommendation technology is to integrate the recommendation system into the travel industry, and through personalized recommendations, it can provide users with assisted decision-making and recommend travel products of interest to them more efficiently and accurately. Therefore, it has a broad application space and research value (Jain, Pamula, and Srivastava 2021).

The content of travel recommendation includes the recommendation of travel products in food, accommodation, transportation, travel, entertainment and shopping. According to the recommendation content, personalized travel recommendation can be roughly divided into three kinds: individual travel product recommendation, travel package recommendation and travel line recommendation (Jain, Pamula, and Srivastava 2021).

Takashita et al (Takashita et al. 2011) used 34,206 travel photos shared on the Flickr platform (Zheng et al. 2010) (uploaded by 1204 users, with an average of 78 photos per place and an average of 15 tags per photo, where tags can be used for indexing or searching. For example, a photo about Longji terraces in Guilin might have the tags: Guilin, Longji, terraces, natural landscape, etc). The system collects metadata of each photo (title, file name, date of

taking the photo, uploader and location information at the time of taking the photo, etc.), user-generated data (tags of the photo), etc. Once the user selects a travel destination, the system recommends to the user a photo set of tourist attractions related to that destination. The similarity between the tags in this photo set is then calculated based on whether the tags in these photo sets are labeled with the same attractions. These tags are grouped and a representative tag is selected for each group, and then these representative tags are also recommended to the user. If the user then selects one or more representative tags, the system then recommends other photos of attractions corresponding to that tag.

Yang et al (Yang et al. 2017) collected multiple travel photos with GPS records on Flickr and clustered the photos into 1,108 categories using the mean drift clustering method. Then the representative photos and labels are calculated for each category. If the user enters a keyword query, the sights represented by the photos corresponding to the representative tags most similar to the keyword are recommended. If the user uses a photo query, the similarity between that photo and each representative photo is calculated, and the attractions corresponding to the representative photo are recommended according to the degree of similarity.

Alhamid et al. (2016) proposed the context rank method. Multiple photos were collected from the website Panoramio (Xu, Chen, and Chen 2015), similar to Yang et al. A mean drift clustering algorithm was used to cluster photos into 120,628 landmarks (referring to unique representative buildings or natural objects in a city, such as the Great Wall in Beijing, the Eiffel Tower in Paris, etc). The popularity of this landmark is estimated based on the number of photos taken by users of the same landmark. And find the most representative photos and tags for each landmark. By analyzing the geographic location information of users' uploaded photos, we use the kernel density estimation method to calculate users' travel geographical preferences. In addition, three other variables are calculated based on the user's travel history (obtained from the user's uploaded photos analysis): visual similarity (the similarity between the visual features of all photos uploaded by the user and the visual features of a representative photo of a landmark), tag similarity (the similarity between the tags of all photos uploaded by the user and the representative tags of a landmark) and collaborative filtering value (the similarity between the user's preference for a landmark using the collaborative filtering method to calculate the user's travel preference value for a landmark). Combining the four values: landmark popularity, visual similarity, tag similarity and collaborative filtering value, the final prediction of the user's preference for a landmark is made using the listNet learning method, and the landmark with the top preference value is recommended to the user.

The route recommendation mainly plans one or more reasonable travel itineraries for users that meet their interests and expectations. Ravi (Ravi et al.

2019) et al. extract tourist attractions and sightseeing spots belonging to the destination from some travel portals according to the user's input, and list them respectively by popularity for users to choose. In this case, sightseeing attractions refer to attractions that have a certain vertical height and can be viewed as a distant view, such as Mount Fuji, Eiffel Tower, Tiananmen Square, etc. After the user selects the starting site, the ending site, a stopover site and the sightseeing spots he wants to see along the way, STIMS (Space-Time Information Management System) is used to calculate the shortest path between the selected sightseeing spots to form multiple candidate tour routes. The average viewability of each tour route to the selected sightseeing spots is then calculated, and the tour routes are recommended to the users in the order of the highest and lowest viewability values.

In traditional personalized recommendation systems, there are four main recommendation methods: content-based recommendation, collaborative filtering-based recommendation and knowledge-based recommendation and combined recommendation. Pessemier, Dhondt, and Martens (2017) gave a short review of recommendation methods in the tourism domain and pointed out the importance of domain knowledge and contextual information for tourism recommendation. Chen et al (Chen et al. 2020) discussed tourism recommendation methods and they suggested that knowledge and session-based methods are more suitable for tourism recommendation. While traditional collaborative filtering methods rely on the user's historical behavior, knowledge-based methods take advantage of the fact that collaterals can express their needs for desired products. For example, the price of the desired hotel and the connection between domain expertise and product attributes, user consumption experience, and user satisfaction recommend a set of products for the user that are as appropriate as possible. The method of the session mainly refers to the type of end use, etc. The method does not suffer from cold starts and does not require a large user base. The two most important factors that affect its performance are knowledge acquisition and interactive user interface.

The framework of travel recommendation system usually contains the following modules: user modeling, user context awareness, travel information base, recommendation engine and presentation of recommendation results. As shown in [Figure 1](#). Among them, user preferences are obtained and modeled through explicit information (survey questionnaire, user rating, etc) and implicit information (web operation, web log mining, etc). Weather information is obtained through a third-party Web service. Obtaining user's contextual information through mobile device perception, etc. Obtaining social context information of users through social networks. The recommendation engine calculates personalized recommendation results for users by mixing several recommendation methods based on the user model and various types of contextual information, feedback from users in the travel knowledge

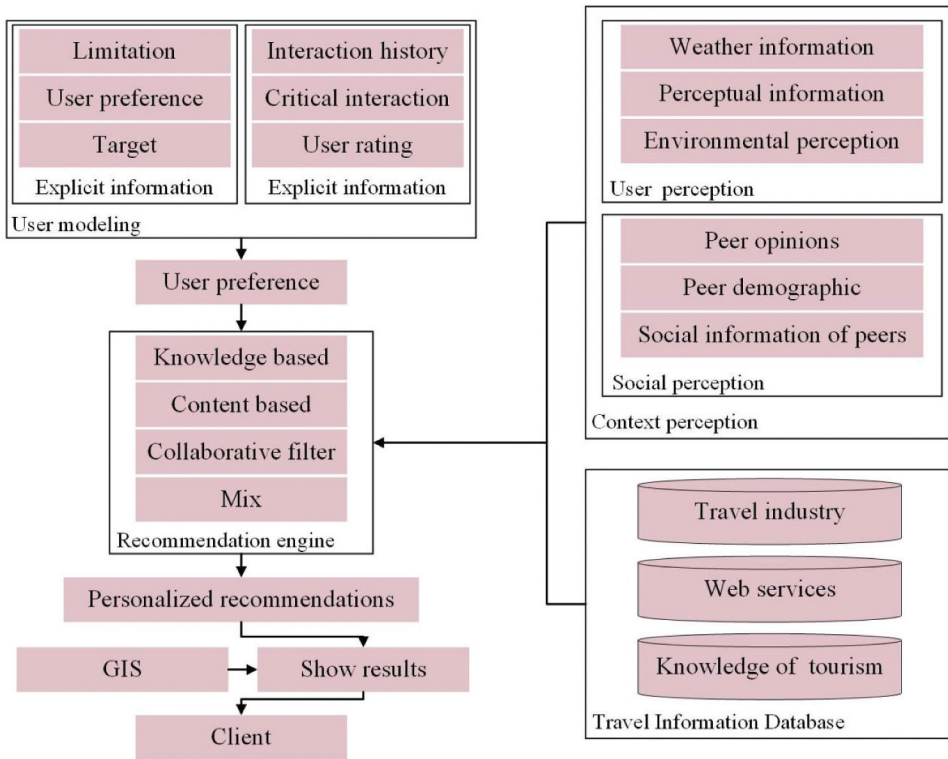


Figure 1. General framework of travel recommendation system.

base, etc. And the results are sorted (Zhu, Yu, and Cai 2019). Finally, a more accurate recommendation list is filtered by combining with GPS positioning system. And the recommendation results are presented to the user in the form of some kind of list or 2D/3D map through the client browser or mobile terminal. Compared with the traditional recommendation framework, the travel recommendation system is more flexible. It not only supports the association combination of multiple basic elements and recommendations to meet the initial user requirements, but also supports diverse human-computer interaction by combining user context perception and GIS map positioning.

Research Status of Collaborative Filtering Recommendation Algorithm

The recommendation techniques used in recommendation systems nowadays mainly include association rule-based recommendation, content-based recommendation, hybrid recommendation, and collaborative filtering. The collaborative filtering algorithm is the most widely used technology nowadays. For example, the daily song list in NetEase Cloud Music, Guess Your Favorite in Taobao Mall, and product recommendation in Jingdong Mall are all based on this algorithm. Simply put, the recommendation system is based on the

user's browsing habits to determine the user's interests. By discovering the user's behavior, we can recommend the appropriate information to the user to meet the user's personalized needs. The structure of a recommendation system usually consists of two parts: online and offline. The online part is the front-end display part. The offline part consists of the back-end logging system and the recommendation algorithm system. The offline part builds a model by learning user profiles and behavior logs, calculates the corresponding recommendation content in a new context, and presents it in the online page (Gu and Liu 2013).

The collaborative filtering algorithm is mainly used to predict the target users' evaluation of similar products by analyzing the users' evaluation of some products, and use it as the basis for making recommendations to users. The collaborative filtering algorithm is a popular and widely used approach in recommendation systems, as it is based on the assumption that users who have similar preferences in the past are likely to have similar preferences in the future. Collaborative filtering can be used to make personalized recommendations by identifying users with similar preferences and recommending items that similar users have liked in the past. In the context of tourism recommendation systems, it is important to provide personalized recommendations tailored to each user's interests and preferences. However, traditional collaborative filtering algorithms may not perform well when there is sparse data or when users have diverse preferences that are not captured by the similarity measures used in the algorithm.

The paper proposes an improved collaborative filtering algorithm based on user preferences to address these limitations. By incorporating user preference information and using the Jeffries-Matusita distance metric, the proposed algorithm aims to improve the accuracy of recommendations and provide more personalized and relevant suggestions to users. The paper argues that incorporating user preferences can improve the accuracy of recommendations by addressing the issue of diversity in user preferences and by taking into account the overall rating behavior of each user. Furthermore, using Jeffries-Matusita distance as a similarity metric allows the algorithm to capture more nuanced differences in user preferences, which traditional similarity measures may not capture.

Using a collaborative filtering algorithm based on user preferences is a suitable choice for this work, as it allows for the development of personalized and accurate recommendations in the context of tourism while addressing the limitations of traditional collaborative filtering algorithms.

The advantages of collaborative filtering (Li and Li 2019) are: 1) good real-time and accuracy of recommendation. 2) no technical barriers. 3) the ability to reduce the impact of data sparsity on the recommendation results.

Two types of collaborative filtering methods have been widely studied, namely Memory-based CF and Model-based CF.

The memory-based collaborative filtering method generally uses the nearest neighbor technique. The distance between users is calculated using their historical preference information. Then the weighted values of the target user's neighboring users' evaluations of the items are used to predict the target user's preferences for a specific item, and the recommendation system makes recommendations to the target user based on the preferences. This method includes User-based CF (Fu, Liu, and Lai 2019) and Item-based CF (Pan et al. 2017). The user-based approach predicts users' ratings based on the similarity of rating behaviors among users. The item-based approach predicts user ratings based on the similarity between the predicted items and the actual items selected by users. Since these two approaches are based on different principles, they perform differently in different application scenarios. The user-based recommendation is more social, reflecting the popularity of items in the user's interest group. Item-based recommendations are more personalized, reflecting the user's own interest heritage. In memory-based collaborative filtering, the common methods for calculating similarity include Pearson correlation coefficient, cosine similarity, Jaccard similarity coefficient, and Euclidean distance (Gasmi et al. 2015). There is no absolute advantage or disadvantage of these similarity calculation methods, and the application scenarios are not absolute. Therefore, they should be chosen flexibly or combined according to the actual application scenarios and data characteristics.

The model-based collaborative filtering approach simulates users' rating behavior for items by modeling them. It uses machine learning and data mining techniques to identify models from training data and use the models to predict the ratings of unknown items. Common models include clustering models (Jo and Rhee 2014), Bayesian models (Zhang et al. 2019, 2022), matrix decomposition (Nguyen and Zhu 2013; Wang, Wang, and Chen 2021), and so on.

The recommendation techniques currently used by websites to meet a variety of user needs include association rule-based recommendations, content recommendations, collaborative filtering, and hybrid recommendation algorithms. Among them, collaborative filtering technology is one of the best technologies with the best performance, which is summarized as follows: 1) The recommendation object covers a wide range. Collaboration-based technology is based on the interests of similar users to generate recommendations. It is not necessary to get information about a specific item and its neighboring items. This allows collaborative filtering techniques to recommend various types of resources to users. 2) Generating new recommendations. Collaborative filtering does not compare the correlation between the specific content of an item and the user's description of that item. The recommended items are not just the user's historical hobbies, which will help in the discovery of the user's potential interests and achieve leapfrog

recommendations. 3) Very low impact on the user. Collaborative filtering only uses ratings as a representation. Users only need to rate some items to get the recommendation service provided by the system. This results in very little user disruption. 4) Low technical difficulty. User preference information is easy to collect and the algorithm procedure is not complicated.

Despite the great success of collaborative filtering algorithms, there are still some problems that limit their development. They are summarized as follows. 1) Sparsity. The actual size of users and projects is huge, and users only rate very few projects, so the data obtained is quite limited and reliability is difficult to be guaranteed. 2) Scalability. The number of users and items in the website is growing continuously, which makes the rating matrix increase rapidly and become a high-dimensional matrix. The complexity of the mathematical calculations designed into the algorithm increases rapidly, making the system slow and users unable to get real-time recommendation content.

Collaborative Filtering Algorithm Based on User Preference

The collaborative filtering algorithm provides personalized recommendations to users by calculating the similarity between users and by nearest neighbor. However, the data in reality is relatively large and it is less likely that different users will all rate an item. The algorithm cannot find the similarity neighbors of a certain user, resulting in a lack of common items between users or locations. Figure 2 shows the general collaborative filtering technique recommendation process.

In fuzzy mathematics, closeness can be used to express the closeness of 2 sets. If the distance between 2 fuzzy sets is larger, the closer the closeness is, the weaker the correlation of the sets. Suppose $u = (u_1, u_2, u_3 \dots u_n)$, then the Hamming closeness is:

$$N(A, B) = 1 - \frac{1}{\sqrt{n}} \left(\sum_{i=1}^n ((A(u_i) - B(u_i))^2)^{\frac{1}{2}} \right) \quad (1)$$

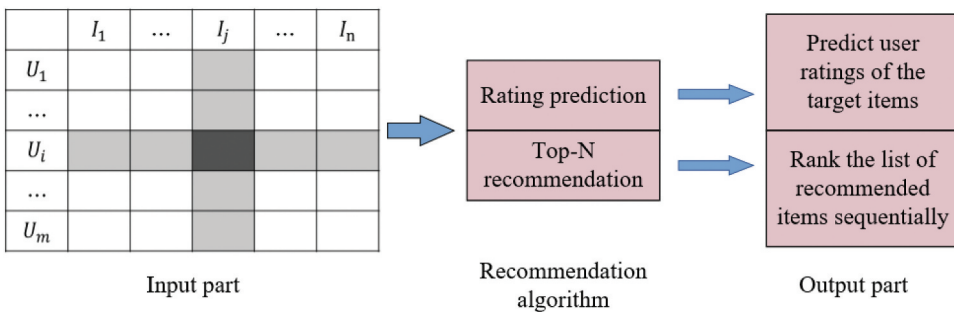


Figure 2. Collaborative filtering technology flow.

The above describes how the concept of “closeness” can be used in fuzzy mathematics to express the similarity between two sets. Specifically, the Hamming closeness is defined as a measure of the similarity between two fuzzy sets A and B. The Equation (1) presented in the text is used to calculate the Hamming closeness between two sets A and B. In the equation, n represents the number of elements in the sets. u_i represents the i th element in the sets A and B, and $A(u_i)$ and $B(u_i)$ represent the membership values of u_i in the sets A and B, respectively. The equation calculates the distance between the two fuzzy sets by computing the square root of the sum of the squared differences between the membership values of each element in the sets. The Hamming closeness is then calculated as 1 minus the normalized distance, which reflects the degree of similarity between the two sets. Overall, the Hamming closeness provides a way to measure the similarity between two fuzzy sets, which can be useful in various applications, including collaborative filtering recommendation algorithms for tourist attractions based on user preferences.

A user’s interest preference can be shown by the item rating. The closer the ratings are to the users, the higher the probability that they will like a particular item. So the closeness algorithm can be used to calculate the preference among users. Defining users’ rating items $L = (l_1, l_2, l_3 \dots l_m)$ and u and v are the set of ratings on L , the closeness based on users’ preferences can be defined as:

$$N(u, v) = 1 - \sqrt{\frac{1}{\sigma_{uv}} \sum_{Iu \cap I_v} (r_{ui} - r_{vi})^2} \tag{2}$$

Equation (2) only uses the common user rating items, which can only reflect the real situation of a small proportion of users, ignoring the user ratings outside the common items and failing to fully grasp the user’s interest preferences. In this paper, we use global rating items to expand the scope of finding users’ interest preferences. Then, considering that the formula calculation on sparse data will be large due to the sparse and uneven distribution of ratings, which will cause distortion of user closeness, we use the idea of stratification to divide the global items into user common items and user local items. For the calculation of the no-common rating part, the average rating is used to fill. The computational expression is:

$$N^{\zeta}(u, v) = \sqrt{\frac{1}{\sigma_{uv}} \sum_{Iu \cap I_v} (r_{ui} - r_{vi})^2 + \sum_{Iv - Iu \cap I_v} r_{ui \cdot \bar{u}} + \sum_{Iv - Iu \cap I_v} r_{vi \cdot \bar{v}}} \tag{3}$$

Each user has their own rating style, some users may not like to give high scores and some users may give the same rating. For the same preferred item, different users may give different values based on their own rating preferences,

which reduces the accuracy of the recommendation. Considering the users' rating items in the global context, the probability distribution of possible ratings for different users is used to handle the user rating preference problem by calculating the probability distribution. The user's rating preference is defined as:

$$P(u, v) = \sqrt{1 - \sum_{\varphi=1}^x \sqrt{\frac{\varphi n}{l(u)} \frac{\varphi m}{l(v)}}} \quad (4)$$

Equation (4) reflects the density of the probability distribution of the global ratings of different users in the discrete interval. Specifically, the Equation (4) presented in the text is used to calculate the similarity score (P) between two users (u and v) based on their global rating information. The equation computes the density of the probability distribution of the global ratings of different users in a discrete interval. In the equation, x represents the number of discrete intervals used to represent the range of possible rating values. n and m represent the number of users who have rated the item being considered at each of the x intervals for users u and v, respectively. l(u) and l(v) represent the total number of items rated by users u and v, respectively.

The equation uses the Hamming approach degree to calculate the similarity between users' global rating distributions, taking into account the density of ratings at each interval. The resulting similarity score reflects the degree of similarity between the two users' overall rating patterns. Overall, Equation (4) provides a method for calculating user similarity based on their global rating distributions, which can be used as a component of the proposed collaborative filtering algorithm for tourist attraction recommendations based on user preferences.

The Jeffries-Matusita (J-M) distance is widely used in various fields such as image processing, signal and pattern recognition. For two components with small distances, poor differentiability is exhibited and poorly similar results are derived from them. In the collaborative filtering algorithm, the user's rating set items can be used to represent the distance components, and eventually the J-M algorithm is used to convert the user's preference in the global item rating into the inter-user similarity calculation. As shown in Equation (5).

$$D(u, v) = \sqrt{1 - e^{-N^{\zeta}(u,v)P(u,v)}} \quad (5)$$

In a recommender system, the number of commonly rated items among different users is highly variable. The more items are rated, the more information is extracted from them, and the more accurate the similarity calculation result is. Therefore, the proportion of the number of co-rated items is a very important influencing factor. In this paper, a logistic function is introduced to

linearly map the common rating items of users to obtain a restrictive penalty factor about the common rating, and the score will be smaller if the users have fewer common items.

The advantage of JMSD algorithm is that it uses the common rating information and absolute rating information of users. However, with sparse data, there may be fewer common items among users, and users lack ratings on items, and the similarity calculation becomes inaccurate. To alleviate this problem, the JMSD algorithm is weighted by combining Equations (4)-(7). The equation is shown in (6).

$$\text{sim}^{\text{JM}}(u, v) = Q_{u,v} \times \text{JMSD} \quad (6)$$

The weighted JMSD algorithm uses the absolute rating information of users and the proportion of common ratings, which is something that the preference degree algorithm does not take into account. Making the combination of user-based preference algorithm and weighted JMSD algorithm can make full use of their advantages. The final approach of this paper is derived as.

$$\text{sim}^{\text{ED-M}}(u, v) = D(u, v) + \text{sim}^{\text{JM}}(u, v) \quad (7)$$

Equation (7) effectively uses the common rating items and all rating information by considering the user's interest preference closeness and the user's rating preferences. This solves the co-rating problem caused by the fact that in sparse data, the parameters for finding similarity in the conventional metric only consider the actual rating information of the co-rated items. Finally, the accuracy of the collaborative filtering algorithm is improved by combining the weighted JMSD algorithm with the ratio of absolute rating information and common ratings of users.

The proposed algorithm model for collaborative filtering recommendations for tourist attractions based on user preferences under sparse data makes several theoretical and managerial contributions.

Theoretical contributions:

The paper proposes an improved algorithm for collaborative filtering that considers user preferences and global and local rating information. The use of Jeffries-Matusita distance as a similarity metric allows the algorithm to capture more nuanced differences in user preferences and improve the accuracy of recommendations.

The proposed algorithm addresses the issue of sparse data by incorporating user preference information and taking into account the overall rating behavior of each user. This provides a more robust and accurate recommendation system that can perform well even with limited data.

The paper provides a theoretical framework for incorporating external data sources and linkages in the preprocessing step to improve the quality and

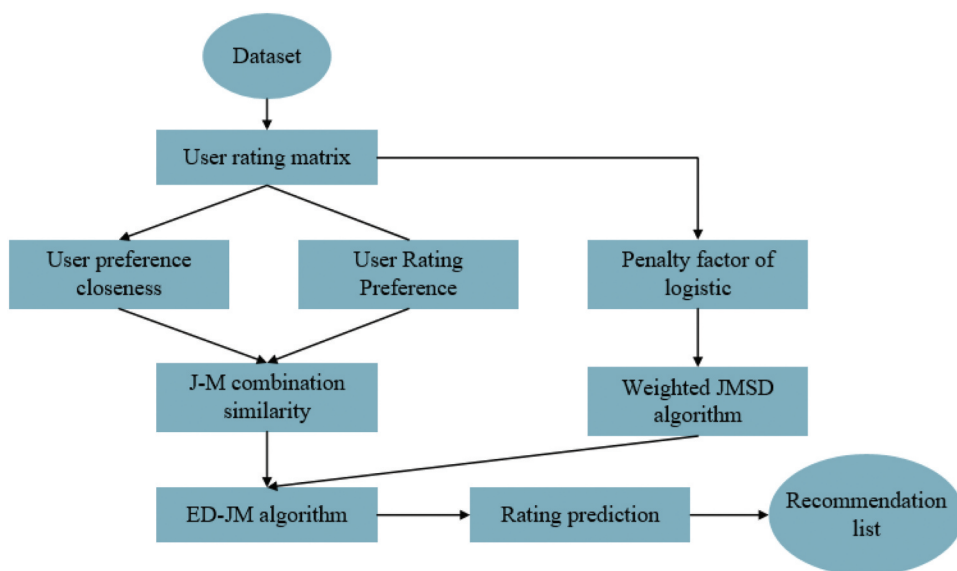


Figure 3. Algorithm flow.

relevance of data and to identify new patterns and insights that were not apparent in the original data sources.

Managerial contributions:

The proposed algorithm has practical implications for the tourism industry by providing a more accurate and personalized recommendation system for tourist attractions. This can improve user satisfaction and engagement, leading to increased revenue and customer loyalty for travel and tourism companies.

Using external data sources and linkages in the preprocessing step can enable travel and tourism companies to capture additional insights and information about user preferences and behavior. This can inform business strategies and marketing efforts and help companies better understand and meet their customers' needs.

The proposed algorithm can be adapted and applied to other recommendation systems in different domains, such as e-commerce, music, or movie recommendations. This has broad implications for the development of personalized recommendation systems that can provide relevant and accurate suggestions to users in various contexts.

The specific algorithm flow is shown in [Figure 3](#).

Table 1. Distribution of experimental data.

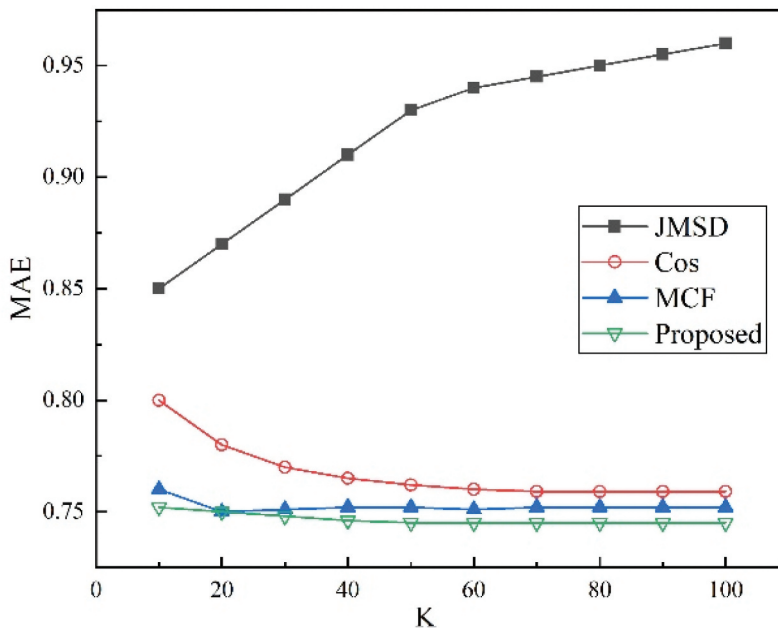
ID	Training set	Test set
Number of users	2933	733
Number of attractions	188	47
Number of ratings	18902	4726
Sparsity grade	0.963	

Experiments

Experiment Preparation

The dataset in this paper is obtained by web crawling information about attractions in a city, including user IDs, attraction IDs and user ratings of the attractions. The dataset includes 45,473 ratings of 286 attractions by 20,354 users, with ratings ranging from 1 to 5. To facilitate the experiments, the obtained data are processed as follows: delete users' duplicate rating records; delete attractions with no user ratings; and delete users with less than 3 rating records. Among them, 80% were used for the training set and 20% for the test set. The final obtained tourism dataset is shown in Table 1.

In terms of evaluation methods, this paper uses the mean absolute error (MAE) and root mean square error (RMSE) to evaluate the algorithm. Normalized discounted cumulative gain (nDCG) is an evaluation metric to measure the accuracy of recommendation lists in information retrieval. In this paper, we use nDCG to test the recommended results. nDCG is a number between (0,1). The larger the value of nDCG, the more accurate the ranking of

**Figure 4.** Comparison of MAE for different similarity measures.

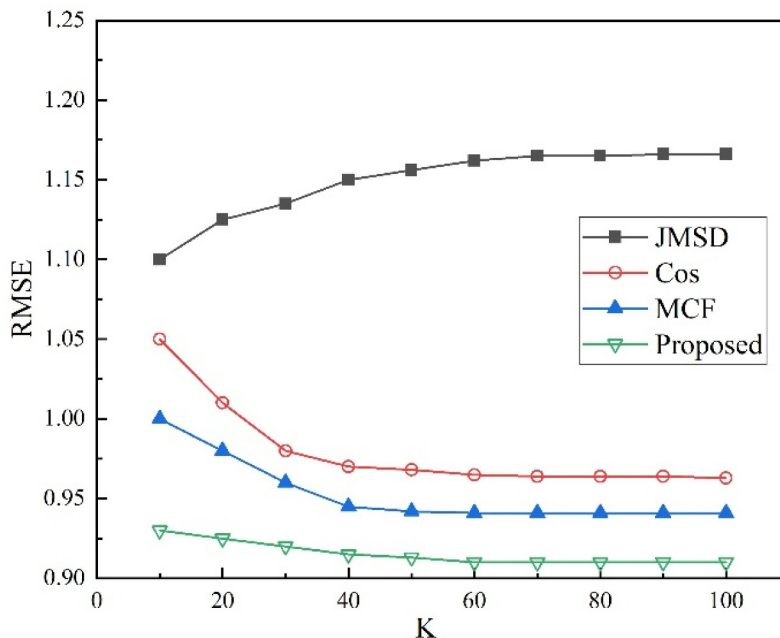


Figure 5. Comparison of RMSE for different similarity measures.

the items in the recommendation list and the higher the accuracy of the recommendation.

Comparison with Traditional Collaborative Filtering Techniques

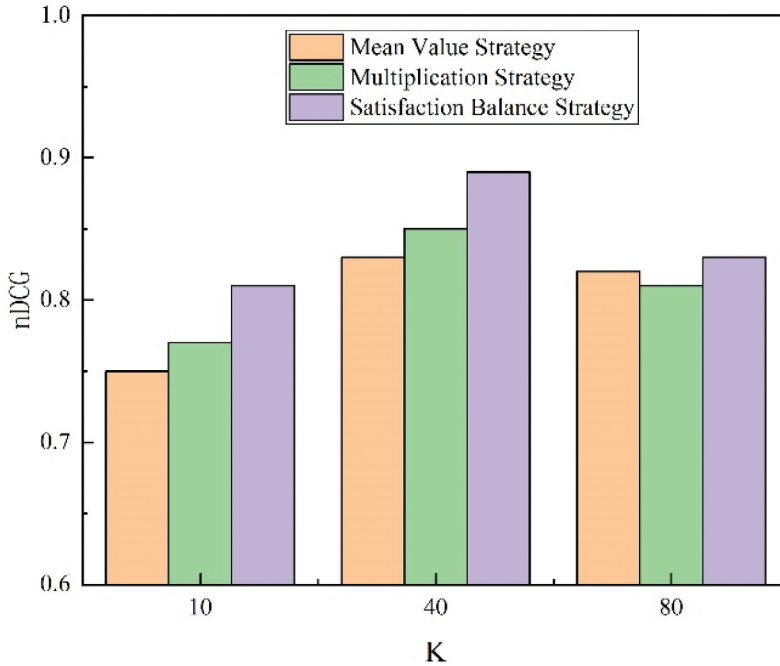
Figures 4 and 5 show the comparison of MAE and RMSE for different similarity measures on the dataset for different values of K. The performance values of MAE, RMSE coverage are calculated based on the values of the neighbor number K, which are defined as values from 10 to 100 with an increment of 10. As can be seen from the Figures 4 and 5, the JMSD algorithm and the modified cosine algorithm are clearly inferior to the other algorithms, except for the above 2 algorithms. The function curve of the algorithm is relatively stable, and the accuracy of the algorithm increases as the value of K increases. The collaborative filtering algorithm proposed in this paper,

Table 2. MAE corresponding to different degrees of similarity.

Number of neighbors	COS	JMSD	MCF	Proposed
10	0.53437	0.51259	0.46796	0.47645
20	0.52441	0.52415	0.46341	0.45244
40	0.52445	0.53879	0.46712	0.44586
60	0.52751	0.5479	0.47239	0.44681
80	0.53124	0.55459	0.47819	0.44961
100	0.53512	0.55954	0.48327	0.45275

Table 3. Algorithm error comparison.

Method	Average MAE	Error Comparison/%	Minimum MAE	Error Comparison/%
COS	0.52894	14.61	0.52365	14.89
JMSD	0.54227	16.72	0.51259	13.05
MCF	0.47235	4.39	0.46341	3.82

**Figure 6.** Impact of the preferred strategy on the algorithm.

which incorporates user preferences, is significantly better than the other methods.

The sparsity of the data may reduce the accuracy of the collaborative filtering algorithm. Table 2 shows the different MAE values for different algorithms on the number of neighbors [10, 20, 40, 80, 100]. Table 3 shows the comparison of the average and minimum error on MAE for the three similarities with the algorithms in this paper. From the data in the table, it can be seen that the algorithm error gradually becomes larger as the value of K. The accuracy of JMSD algorithm is related to the number of user items, so the algorithm still has a poor MAE compared to other algorithms. traditional collaborative filtering algorithm COS only considers user common items, so the overall performance is poor under sparse data, as in Table 3, the lowest error rate of cosine algorithm reaches 14.89%.

In addition, this paper also verifies the effect of different preferences on the performance of the algorithm. The mean strategy, multiplication strategy and satisfaction balance strategy are verified in this paper based on the number of

neighbor distance of 10, 40 and 80, respectively. The experimental results are shown in [Figure 6](#), using nDCG as the evaluation index. The results show that the collaborative filtering algorithm incorporating the satisfaction balance strategy works better.

Discussion – Conclusion

Travel recommendation is a specific application area of intelligent recommendation systems, and collaborative filtering recommendation technique is the most commonly used recommendation algorithm. For the sparsity problem of collaborative filtering in recommendation, this paper proposes an improved collaborative filtering algorithm that incorporates user preferences. Under sparse data, the lack of common rating items may cause the similarity calculation among users to be incomplete, which will reduce the accuracy of travel place recommendation. To address the shortcomings of the traditional collaborative filtering algorithm, this paper first considers users' interest preferences on tourist attractions, introduces the concept of closeness and fuses global rating and local rating to calculate the preference degree among users. The scope of similarity finding is expanded and the effect of sparsity is reduced. Experiments on the classical dataset of tourism show that the recommendation effect of collaborative filtering technique incorporating user preferences is better.

The proposed algorithm model for collaborative filtering recommendations for tourist attractions based on user preferences under sparse data has shown promising results in improving the accuracy of recommendation systems in the tourism domain. Several potential future directions could be explored based on the paper's findings.

First, the proposed algorithm model could be further evaluated and tested on larger and more diverse datasets to determine its scalability and robustness. This would help to establish the generalizability of the proposed approach across different tourism domains and user populations.

Second, the paper could be extended to incorporate other types of user data, such as social media activities, geolocation, and search history, which may provide additional insights into user preferences and improve the accuracy of recommendations.

Third, the paper could explore other machine learning techniques, such as deep learning, reinforcement learning, or hybrid models, to improve the performance of the recommendation system further.

Also, data preprocessing is an important step in any machine learning or data analysis project, including recommendation systems. It involves transforming raw data into a format that is suitable for analysis and modeling, and includes tasks such as cleaning, normalization, feature selection, and dimensionality reduction. Proper data preprocessing can increase the novelty of the

work by improving the quality and relevance of the data used for modeling and analysis, and by identifying new patterns and insights that were not apparent in the raw data.

Linking the data used in a recommendation system to external data sources can further increase the novelty of the work. For example, incorporating data from social media platforms, travel review websites, or location-based services can provide additional information about user preferences and behavior, which can be used to improve the accuracy and relevance of recommendations. Moreover, linking data to external sources can enable the discovery of new patterns and insights that were not apparent from the original data sources, thereby increasing the novelty of the work. In addition, incorporating diverse data sources in the preprocessing step can help to address the issue of sparse data and increase the coverage of the recommendation system. By leveraging data from different sources, the system can better capture the diversity of user preferences and improve the accuracy and relevance of recommendations for a wider range of users.

There are several potential advantages and disadvantages of the proposed collaborative filtering algorithm for tourist attractions based on user preferences.

Advantages:

- (1) Incorporating user preferences: The proposed algorithm considers the user preferences, which can improve the accuracy of recommendations and increase user satisfaction. This is because it considers the user's interests, past behaviors, and ratings when making recommendations.
- (2) Addressing sparse data: The proposed algorithm uses a similarity algorithm based on Jeffries-Matusita distance to address the issue of sparse data and the lack of common rating items among users that traditional similarity measures cannot calculate. This can potentially improve the accuracy of recommendations on sparse datasets.
- (3) Improved performance: According to the experimental results presented in the paper, the proposed algorithm outperforms traditional collaborative filtering algorithms and other algorithms, achieving high accuracy even on sparse tourism data sets.

Disadvantages:

- (1) Complexity: The proposed algorithm may be more complex than traditional collaborative filtering algorithms, as it involves calculating user preferences using global and local rating information and using Jeffries-Matusita distance to measure similarity. This could make the algorithm more difficult to implement and require more computational resources.

- (2) Data preprocessing: The proposed algorithm may require extensive preprocessing of data to calculate user preferences and other necessary information. This could be time-consuming and may require expertise in data processing and analysis.
- (3) Limited applicability: The proposed algorithm may not apply to allegation systems or datasets. For example, it may not be effective in recommending new or niche tourist attractions with limited rating data.

Overall, the proposed algorithm shows promise in improving the accuracy and user satisfaction of tourist attraction recommendation systems by incorporating user preferences and addressing the issue of sparse data. However, its complexity and data preprocessing requirements may limit its applicability in some cases.

Finally, the paper could be extended to investigate the effectiveness of the proposed algorithm model in real-world scenarios, such as in online travel booking platforms or mobile tourism applications. This would help to determine the practical value of the proposed approach in the tourism industry and its impact on user satisfaction and engagement.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This paper was supported by the Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (No.2019SJA0969).

Data Availability Statement

The data used in the research are available from the author upon request.

References

- Alhamid, M. F., M. Rawashdeh, H. Dong, M. A. Hossain, and A. El Saddik. 2016. Exploring latent preferences for context-aware personalized recommendation systems. *IEEE Transactions on Human-Machine Systems* 46 (4):615–23. doi:10.1109/THMS.2015.2509965.
- Bai, M. L., R. Pamula, and P. Kumar Jain. 2019. Tourist recommender system using hybrid filtering. In *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, 746–49. doi:10.1109/ISCON47742.2019.9036308.
- Chen, C., J. Jiang, and R. Fu. 2021. An intelligent caching strategy considering time-space characteristics in vehicular named data networks. *IEEE Transactions on Intelligent Transportation Systems*.

- Chen, C., J. Jiang, Y. Zhou, N. Lv, X. Liang, and S. Wan. 2022. An edge intelligence empowered flooding process prediction using internet of things in smart city. *Journal of Parallel and Distributed Computing* 165:66–78. doi:10.1016/j.jpdc.2022.03.010.
- Chen, L., Z. Wu, J. Cao, G. Zhu, and Y. Ge. 2020. Travel recommendation via fusing multi-auxiliary information into matrix factorization[J]. *ACM Transactions on Intelligent Systems and Technology (TIST)* 11 (2):1–24. doi:10.1145/3372118.
- Chen, C., Y. Zeng, and H. Li. 2022. A multi-hop task offloading decision model in MEC-Enabled internet of vehicles. *IEEE Internet of Things Journal*.
- Cui, G., J. Luo, and X. Wang. 2018. Personalized travel route recommendation using collaborative filtering based on GPS trajectories. *International Journal of Digital Earth* 11 (3):284–307. doi:10.1080/17538947.2017.1326535.
- Du, J., C. Jiang, Z. Han, H. Zhang, S. Mumtaz, and Y. Ren. 2019. Contract mechanism and performance analysis for data transaction in mobile social networks. *IEEE Transactions on Network Science and Engineering* 6 (2):103–15. doi:10.1109/TNSE.2017.2787746.
- Floegel, D. 2020. Labor, classification and productions of culture on Netflix[J]. *Journal of Documentation* 77 (1):209–28. doi:10.1108/JD-06-2020-0108.
- Fu, W., J. Liu, and Y. Lai. 2019. Collaborative filtering recommendation algorithm towards intelligent community. *Discrete & Continuous Dynamical Systems-S* 12 (4 & 5):811. doi:10.3934/dcdss.2019054.
- Gasmi, I., H. Seridi-Bouchelaghem, L. Hocine, and B. Abdelkarim. 2015. Collaborative filtering recommendation based on dynamic changes of user interest. *Intelligent Decision Technologies* 9 (3):271–81. doi:10.3233/IDT-140221.
- Gu, J. S., and Z. Liu. 2013. *The improved collaborative filtering recommendation algorithm based on cloud model[C]//Applied mechanics and materials*, vol. 411. Trans Tech Publications Ltd.
- Hao, Y., and N. Song. 2021. Dynamic modeling and analysis of multidimensional hybrid recommendation algorithm in tourism itinerary planning under the background of big data. *Discrete Dynamics in Nature and Society*.
- Jain, P. K., R. Pamula, and G. Srivastava. 2021. A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review* 41 (August):100413. doi:10.1016/j.cosrev.2021.100413.
- Jo, H. J., and P. K. Rhee. 2014. Distributed recommendation system using clustering-based collaborative filtering algorithm. *The Journal of the Institute of Internet, Broadcasting and Communication* 14 (1):101–07. doi:10.7236/JIIBC.2014.14.1.101.
- Li, X., and D. Li. 2019. An improved collaborative filtering recommendation algorithm and recommendation strategy[J]. *Mobile Information Systems*.
- Lin, S., B. Zheng, G. C. Alexandropoulos, M. Wen, F. Chen, and S. Mumtaz. 2020. Adaptive transmission for reconfigurable intelligent surface-assisted OFDM wireless communications. *IEEE Journal on Selected Areas in Communications* 38 (11):2653–65. doi:10.1109/JSAC.2020.3007038.
- Memon, I., L. Chen, A. Majid, M. Lv, I. Hussain, and G. Chen. 2015. Travel recommendation using geo-tagged photos in social media for tourist. *Wireless Personal Communications* 80 (4):1347–62. doi:10.1007/s11277-014-2082-7.
- Nguyen, J., and M. Zhu. 2013. Content-boosted matrix factorization techniques for recommender systems. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 6 (4):286–301. doi:10.1002/sam.11184.
- Pan, X., W. Zhou, Y. Lu, and R. Li. 2017. User collaborative filtering recommendation algorithm based on adaptive parametric optimisation SSPSO. *International Journal of Computing Science and Mathematics* 8 (6):580–92. doi:10.1504/IJCSM.2017.088977.

- Pessemier, T. D., J. Dhondt, and L. Martens. 2017. Hybrid group recommendations for a travel service. *Multimedia Tools and Applications* 76 (2):2787–811. doi:10.1007/s11042-016-3265-x.
- Ravi, L., V. Subramaniaswamy, V. Vijayakumar, S. Chen, A. Karmel, and M. Devarajan. 2019. Hybrid location-based recommender system for mobility and travel planning. *Mobile Networks and Applications* 24 (4):1226–39. doi:10.1007/s11036-019-01260-4.
- Smith, B., and G. Linden. 2017. Two decades of recommender systems at Amazon. Com[J]. *IEEE Internet Computing* 21 (3):12–18. doi:10.1109/MIC.2017.72.
- Takashita, T., Y. Abe, T. Itokawa, T. Kitasuka, and M. Aritsugi. 2011. Design and implementation of a system for finding appropriate tags to photos in Flickr from web browsing behaviour. *International Journal of Web and Grid Services* 7 (1):75–90. doi:10.1504/IJWGS.2011.038385.
- Wang, X., C. Wang, and J. Chen. 2021. Optimized gradient descent and pre-filled for svd recommendation algorithm. *Journal of Nonlinear and Convex Analysis* 22 (10):2073–88.
- Xu, Z., L. Chen, and G. Chen. 2015. Topic based context-aware travel recommendation method exploiting geotagged photos. *Neurocomputing* 155:99–107. doi:10.1016/j.neucom.2014.12.043.
- Yang, L., L. Wu, Y. Liu, and C. Kang. 2017. Quantifying tourist behavior patterns by travel motifs and geo-tagged photos from Flickr. *ISPRS International Journal of Geo-Information* 6 (11):345. doi:10.3390/ijgi6110345.
- Zhang, C., X. Duan, F. Liu, X. Li, and S. Liu. 2022. Three-Way Naive Bayesian collaborative filtering recommendation model for smart city[J]. *Sustainable Cities and Society* 76:103373. doi:10.1016/j.scs.2021.103373.
- Zhang, T., W. Li, L. Wang, and J. Yang. 2020. Social recommendation algorithm based on stochastic gradient matrix decomposition in social network. *Journal of Ambient Intelligence and Humanized Computing* 11 (2):601–08. doi:10.1007/s12652-018-1167-7.
- Zhang, Y., K. Meng, W. Kong, Z. Y. Dong, and F. Qian. 2019. Bayesian hybrid collaborative filtering-based residential electricity plan recommender system. *IEEE Transactions on Industrial Informatics* 15 (8):4731–41. doi:10.1109/TII.2019.2917318.
- Zheng, N., Q. Li, S. Liao, and L. Zhang. 2010. Which photo groups should I choose? A comparative study of recommendation algorithms in Flickr. *Journal of Information Science* 36 (6):733–50. doi:10.1177/0165551510386164.
- Zhu, L., L. Yu, and Z. Cai. 2019. Toward pattern and preference-aware travel route recommendation over location-based social networks. *Journal of Information Science & Engineering* 35:5.