

INVESTMENT DECISION OF TOURISM LEISURE PROJECT BASED ON COLLABORATIVE FILTERING ALGORITHM

B.H. XUE[†] and T. LIU[‡]

[†] *Institute of Physical Education, Anhui Polytechnic University, Wuhu, China.*

[‡] *Wannan Medical College, Wuhu, China.
RalapMSWU@yahoo.com*

Abstract— In the era of the Big Bang, users have to spend a lot of time looking for the information they really need, and the search engine can't present information that is not described by the user. Based on the User based collaborative filtering recommendation algorithm and the collaborative filtering recommendation algorithm, this paper proposes User-CF algorithm, User-CF-1 algorithm, Item-CF algorithm, Item-CF-1 algorithm, and finally integrates four algorithms to obtain the cooperative strategy based on collaborative filtering is a hybrid recommendation algorithm, namely the Final algorithm. It has been verified that the Final algorithm can effectively solve the long tail problem in the travel strategy recommendation and greatly increase the coverage of the recommended strategy.

Keywords—DIY model; rural tourism; recommended algorithm.

I. INTRODUCTION

In today's era, the term "network" is no longer strange. It can even be said that everyone is dealing with the Internet every day, and the Internet has penetrated into every aspect of our lives (Varela-Veiga *et al*, 2017). People watch movies and TV shows on video sites such as Fantastic Art and Mango TV, and they watch news and current events on portals such as Sina Weibo and Sohu News. People select a wide variety of products on shopping sites such as Taobao and Jingdong Mall, and they can enjoy the sound of nature in QQ music, Baidu music and other music websites... The birth of the Internet has brought great convenience to people (Prince, 2017). At the same time, huge amounts of information have swarmed and made us dazzled and confused. People have already entered the era of information explosion (Hu *et al*, 2017). Users have to spend a lot of energy to find the information they really need. However, there are too many useless information (Siow *et al*, 2015). There are actually too many things. How can they choose the best information in the complex information and how to effectively present the information? In the face of users, this undoubtedly brings new challenges to computer workers (Carrillo *et al*, 2015). To solve this problem, search engines came into being. The search bar can be found on various websites (Peng *et al*, 2016). If the user clearly identifies the product to be searched for, it can be searched accurately. However, most users do not have a clear

search target in most cases (Luo *et al.*, 2017). The search engine cannot present information that is not described by the user. For example, it is not easy to describe the favorite film and television drama in a simple text form (Yong Qin *et al.* 2018).

II. STATE OF THE ART

The recommendation system is applied in various fields. Its application range is distributed in several kinds of recommendations, such as news, book, music, electrical appliance, film and television, web page and other fields (Akyildiz *et al.*, 2018). In the field of e-commerce, personalized recommendation services play a huge role in the Amazon platform. By counting the users who consume on the website, it is only 16% who know exactly what they want to buy, thereby increasing certain economic benefits. In the field of news recommendation, Group Lens is currently a well-known personalized news recommendation system. By analyzing the scores and other records that users make on the news they are hunting, the system infers which news the users like and the selected match. The user's reading habit is recommended for news (Sarma *et al.* 2014). In the film and television industry, Netflix is the largest online movie rental company in the United States and the first to use a recommendation system. According to the website's statistics, 60% of users find a favorite movie CD through a recommendation system. Hulu is an online video site abroad. Since the use of the recommendation system, the CTR has increased by more than 10%. There are many other examples of application recommendation systems, such as MovieLens in the movie field, Pandora in the music field, and You Tube in the video field. In the web page recommendation system, Fab implements targeted recommendations for each user. There is also a personalized portal My Yahoo added to the yahoo web site.

III. METHODOLOGY

A. User-based Collaborative Filtering Recommendation Algorithm

The similarity between items should be calculated. The user's recommendation list is generated based on the user's historical behavior and the similarity between the calculated items. The title of the Amazon website's recommended item is "Customers who viewed this

product also viewed these items" Starting from this definition, the similarity of the item can be defined by the following formula:

$$w_{ij} = \frac{N(i) \cap N(j)}{|N(i)|} \quad (1)$$

In Equation (1), denominator $N(i)$ is used to indicate how many users like item i , and numerator $N(i) \cap N(j)$ is used to indicate how many users are at the same time like i, j . Therefore, the above Eq. (1) can be used to understand the proportion of users who are interested in item j in the user who is interested in item i . The Eq. (1) seems plausible, but there are still problems. For example, if item j is quite popular and most people are interested, then the closer w is to 1. Therefore, the use of this equation will most likely result in a situation where there is a large degree of similarity between any item and a popular item. This is obviously not a good feature for the recommendation system. In order to reduce the recommendation of only popular products to users, the following equation can be used:

$$w_{ij} = \frac{|N(i) \cap N(j)|}{\sqrt{|N(i)| |N(j)|}} \quad (2)$$

The Eq. (2) achieves the desired goal by penalizing the weight of item j , avoiding the possibility of getting many items similar to hot items. From the above definition, it can be seen that in the item based collaborative filtering algorithm, because many users are interested in these two items, the two items will generate similarity, that is, any user's historical behavior list can contribute to item similarity. There is a hypothetical implication here, that is, each user's preferences are limited to a few aspects, and so if there are two items in the same user's interest list, then they may belong to a limited number of areas, so there is a lot of similarity. The algorithm of this paper is similar to the user-based collaborative filtering recommendation algorithm. When using it, we must first create a user-inverted list to calculate the similarity between items.

After that, for each user, two items in the item list must be created. Two are added to the co-occurrence matrix C . Fig. 1 is a small example of calculating similarity between items. The table in the left side of the figure with five rows and one column is the behavior record of the user, and each row represents a collection of items that likes a user. First of all, we will increase the number of items in each item collection by two. Then, the obtained five matrices are added to obtain the matrix C described above. Among them, $C[i][j]$ is the number of users who are interested in both item i and item j . Finally, the matrix C is normalized to obtain the cosine similarity matrix W between articles. After calculating similarity between items, this paper uses the following equation to calculate how much user u likes item j :

$$P_{uj} = \sum_{i \in N(u) \cap S(j, K)} w_{ji} r_{ui} \quad (3)$$

In the above equation, $N(u)$ denotes all the items the

user u likes. $S(i, K)$ denotes the K items most similar to the item i , and w_{ji} denotes the similarity between the item j and the item i . Degree r_{ui} is used to indicate to what degree the user u likes item i . Due to the use of a single behavior of implicit feedback data, if the target user u has ever acted on it for item i , then we assume that $r_{ui}=1$. Eqn. (3) means that there are some items that are more similar to the favorite items in the historical behavior of the target user, and then their ranking in the user's recommendation list will be higher.

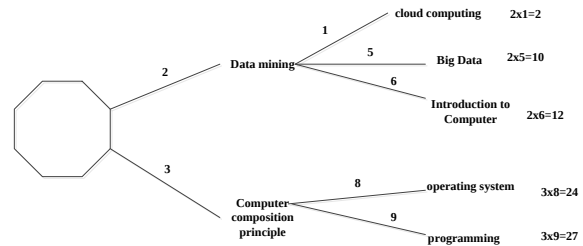


Figure 1. User based collaborative filtering algorithm

Fig. 1 is a simple example of this recommendation algorithm: "Data Mining" and "Principle of Computer Composition" are two books that the target user definitely likes. The algorithm first finds the three books that are most similar to the two selected books, then it calculates the target users' preference for each book according to the formula. For example, this algorithm recommends the book "Introduction to Computers" to users because it is similar to "Data Mining" and has a similarity of 1, which is similar to the "Principles of Computer Composition" and has a similarity of 2. The user's interest in the book "Data Mining" is 5, and the interest in the book "Principles of Computer Composition" is 3, so the user's interest in the "Introduction to the Computer" is $1 \times 2 \times 3 \times 5 = 30$.

B. User-based collaborative filtering travel strategy recommendation algorithm

This paper introduces the traditional user-based collaborative filtering recommendation algorithm. It will slightly modify this algorithm and apply it in the process of the tourism strategy. In this paper, we call this algorithm User-CF algorithm. Recommended cities calculate user ratings based on user behavior history. Each user has a different behavior history for multiple raiders. This behavior can be liked, browsed, or created.

Each raider has a corresponding tourist city, and different behaviors corresponds to the user's different ratings for the city. Therefore, calculate a user's scoring algorithm for a city as: take the user's behavior rating for all the raiders that include the city. Calculate the similarity between users and users according to the user's score on the city, as follows:

$$w_{ij} = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)| |N(v)|}} \quad (4)$$

In Eqn. (4), $N(u)$ represents the set of cities and ratings

included by user u , and $N(v)$ represents the set of cities and ratings included by user v . Calculate the city recommendation degree for the user's recommended city as follows:

$$P(u, i) = \sum_{v \in S(u, K) \cap N(i)} w_{uv} r_{vi} \quad (5)$$

In Eqn. (5), w_{uv} denotes the similarity of user u and user v , and r_{vi} denotes the score of user v for city i . $S(u, K)$ denotes the set of K users with the highest similarity with user u . $N(i)$ represents the set of city scores of the K users with the highest similarity to user u . According to the city recommended raiders, calculate the city's score on the Raiders. Unlike a user who has a history of behavior in the city, a city scores only 0 and 1 for a strategy. If a strategy contains a city, the city adds 1 to the score of the strategy to calculate the similarity between the cities. (4) Calculate the recommended degree of raiders for city recommended raiders, using Eqn. (5). Integrating the city recommendation degree of the chosen city in the first and second steps of the user and the selected recommendation level of the city recommended strategy, calculate the mentioned level of the user recommended strategy. This step is the unique process of the User-CF algorithm applied to the tourism strategy recommendation system, and it belongs to the improvement of the algorithm. The city recommendation degree and the strategy recommendation degree obtained from the previous two steps are finally recommended to the user to get the travel strategy as follows:

$$P(u, b) = \sum_{c \in Rc, b \in Rb} r_{uc} r_{cb} \quad (6)$$

Among them, $P(u, b)$ denotes the recommendation degree of the recommended raiders b to the user u , and the right of the equal sign of the Eqn. (6) Rc below the sign indicates the set of cities recommended to the user u , while the lower Rb indicates the city c recommended Raiders set, equal to the right before the sign of the r . r_{uc} denotes recommending the city recommendation degree of the city c to the user u , and r_{cb} after the multiplication number represents the recommendation degree of the recommended strategy for the city c . Calculate recommended travel guide coverage. Using the traditional method of calculating coverage, follows:

$$\text{Coverage} = \frac{|U_{u \in U} R(u)|}{|I|} \quad (7)$$

In Eqn. (7), $R(u)$ denotes a set of Raiders recommended to user u , and $|U_{u \in U} R(u)|$ denotes the number of Raiders recommended to all users. The $|I|$ denotes all Raiders.

The textual representation of the coverage rate is the proportion of all items selected through the referral system to the total items on the referring website. For example, assumptions can be made assuming that a set of all users of a website is K , and that a length N item selected by using the recommendation system of the website is recommended to each user's article list $R(k)$ It

can be seen that Coverage is an evaluation indicator that the supplier of the item is very fancy. Like the cosmetics recommendation platform, the manufacturer will be very concerned about whether their cosmetics are recommended by the sales site. If the coverage of a recommended system is high, then the probability of those unpopular items being recommended out is greatly increased. In addition, the recommended coverage rate of popular products is low, because if only popular products are recommended, the number of hot items is a small number and the proportion of total items is small. A good recommendation system is not simply to satisfy users, but also to increase coverage.

In summary, in theory, the two improved algorithms selected in this paper can solve the problems raised in this study very well, but the specific application effect needs to be further verified.

IV. RESULTS AAND DISCUSSION

Because the improvement is based on the recommendation algorithm of the tourism strategy, the experiment needs to analyze a large amount of data on the travel website. It needs to collect the user's historical behavior record and the content of the tourism strategy. System management data includes input data, model data, and output data. The input data includes user information, travel guide information, user rating information, and the like. The user information is obtained from the personal information filled in after the user logs in to the system, including the user's identity, gender, age, occupation, address, and email address. Travel guide information includes travel locations, tourist cities, listing dates, types of play, and tourism projects. The user rating information refers to the user's assessment of the travel strategy. There are various forms. It may be a direct score, a written comment, or a fuzzy evaluation.

This data uses the user's method of scoring travel tips. Model data includes model input data and model output data. Different algorithms correspond to different input data. Therefore, in the "data preprocessing" phase, the input data is processed as model input data, user data, raiders data, scoring data. Model structure data and user classification data are two parts of the model output data, while the classification results and classification results of the original users are user classification data. The output data includes user-predicted rating data, predicted new-raised user data, and predicted new user rating data. The experimental data set of this article was collected from the real users, real behaviors, and real data of the travel website. There was a total of 42,523 historical records of user behavior, 1,452 cities, and 10,389 high-quality strategies. Experiment A was as follows: The number of users recommended by the User-CF algorithm was 265, and the coverage was 8.25%. The number of recommendations to the city was 6445, and the coverage was 62.0%. The number of user-recommended raiders is 774 and the coverage is 7.45%. Using the improved User-CF algorithm in this

experiment, the number of recommended cities for the user is 623, and the coverage is 42.91%. The number of recommended cities for the city is 6,109, and the coverage is 58.8%. The number of users recommended raiders is 1932, and the coverage is 18.6%. It can be seen that the coverage of the improved User-CF-1 algorithm is greater than the coverage of the underlying User-CF algorithm. The coverage table and comparison chart are shown in Table 1 and Fig. 2:

Table 1. Improved coverage before and after

	user-city coverage	city-book coverage	user-book coverage
User-CF	10%	35%	11%
User-CF-1	28%	25%	23%

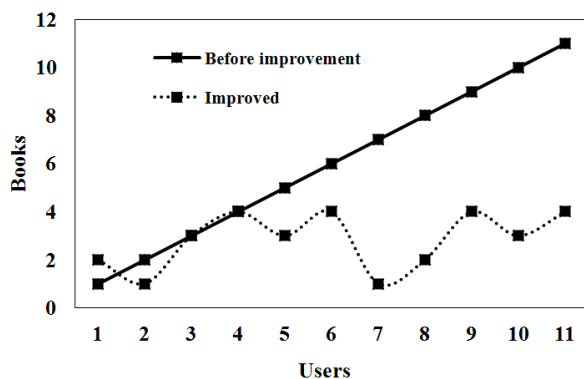


Figure 2. User improved algorithm before and after.

Experiment B was as follows. The number of cities recommended to the user using the tem-CF algorithm was 696, the coverage was 7.93%, and the number of recommended cities for the city was 6,201, and the coverage was 59.69%. The number of recommendations to the user is 2279, and the coverage is 21.94%. The number of cities recommended to the user using the improved Item-CF algorithm in this experiment is 996, and the coverage is 68.6%. The number of recommended cities for the city is 6,201, and the coverage is 59.69%. The number of user recommended raiders is 2762, and the coverage is 26.59%. It can be seen that the coverage of the improved Item-CF algorithm is greater than the coverage of the underlying Item-CF algorithm. Its coverage table and comparison chart are shown in Table 2 and Fig. 3:

Table 2. Steamed dumplings and steamed bread

	vegetables	Meat diet	Curry rice
Item1	13%	56%	78%
Item2	21%	43%	89%

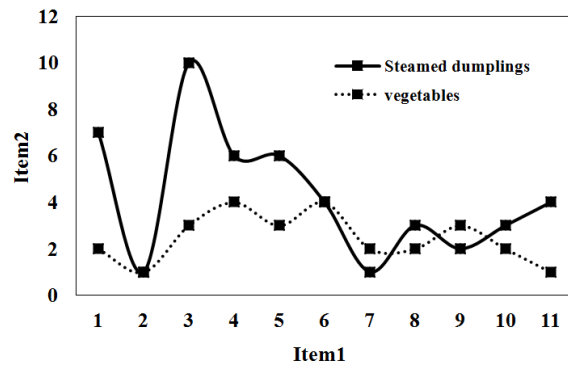


Figure 3. Comparison of coverage.

Experiment C was as follows: The number of recommended raiders for the user obtained using the User-CF algorithm was 774, and the coverage was 7.45%. The number of recommended raiders given to the user using the user-based improved User-CF1 algorithm is 1932, and the coverage is 18.6%. The number of recommended Raiders given to users using the TEM-CF algorithm is 2279, and the coverage is 21.94%. The use of the item-based improved tem-CF1 algorithm to give the user the number of recommended raiders is 2762, and the coverage is 26.59%. The number of recommended raiders given to the user using the hybrid recommendation final algorithm is 4291, and the coverage is 41.3%. The recommended recommendation of the hybrid recommendation algorithm is excellent. An algorithm coverage table and comparison chart is shown in Fig. 4:

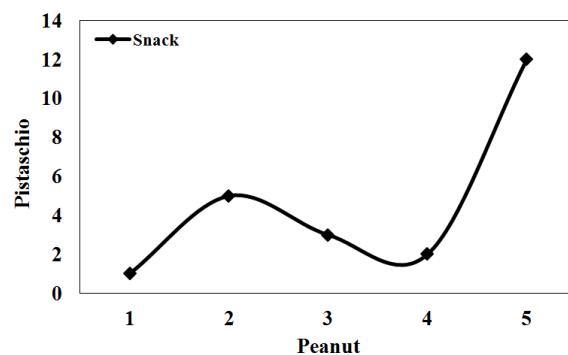


Figure 4. Chestnut and cashew nuts.

Through the above three sets of experiments and analysis results, we can see that each algorithm's improvement has a certain degree of coverage on the basis of the original. The final hybrid recommendation algorithm improves the coverage of nearly 35% compared to the basic algorithm, and the relative effect better. In this example, the following example is used to illustrate that the algorithm can recommend a low-level cold outdated raider to the user. Input a user id, for example, 40, output the recommended raiders and raiders heat, and the first line of the obtained result is [8204, 597. 5], [8205, 264. 0], [21, 155. 5], [6161, 141.5], [8230, 119. 0], [47, 104.5], [8225, 98. 0], [41, 97. 0], [6169, 84. 0], [4130, 82. 0], and the second line

is [2665, 4. 5], [5684, 4. 0], [8447, 4. 0] , [7648, 3.5], [3324, 3.5], [1416, 3.5], [9437, 3.5], [1670, 3.5], [1665, 3. 0], [6541, 3. 0]]. We can see the first behavior of the algorithm to recommend this user popular raiders and raider id, and the second act to the user to recommend unpopular raiders and raider id, where you can explain the algorithm can recommend hot or unpopular raiders to users.

V.CONCLUSIONS

In this paper, the user-based collaborative filtering algorithm is slightly improved to get the User-CF algorithm, and on this basis by improving the degree of recommendation formula, User-CF-1 algorithm can be improved. This paper tries to improve collaborative filtering based on the article. The algorithm is recommended to obtain the Item-CF algorithm. Based on the algorithm, an improved Item-CF-1 algorithm is obtained. Finally, the four algorithms are integrated to obtain a tourism strategy recommendation algorithm based on collaborative filtering. The recommended algorithm for tourism strategy, which can solve the long tail problem, greatly improves the coverage of the hybrid recommendation algorithm, namely the Final algorithm. The collaborative recommendation algorithm based on collaborative filtering is used to recommend travel strategies. Although it is already perfect, it still has its shortcomings. From the data point of view, the amount of data is insufficient. In fact, quality users of high-quality Raiders are only a small part.

ACKNOWLEDGEMENTS

The authors acknowledge Anhui Provincial-level University leading talent introduction and cultivation project. (gxyqZD2016008) .

Anhui Tourism Young expert cultivation project. (AHLYZJ201517)

Research on Anhui Auto-Camping Industry Boosting System and its development Mode. (Grant of Anhui Sports bureau) (ASS2015309)

Study on the cultivation mode of Innovative College talents on experiential education. (2015jyxm170)

REFERENCES

- Carrillo, I.M. and E.M.C. Pérez, “The Design of a Methodology to Develop Competitive Rural Tourism Products and their Implementation in the Municipalities of Colima and Comala, Mexico”, *Global Marketplace*, 138-145 (2015).
- Hu, J., S. Chang, X. Li et al., “Discussion on the DIY-training mode of Undergraduate in the Wood Science and Engineering Specialty”, *Furniture & Interior Design*, **167(7)**, 788-817 (2017).
- Luo, X., L. Peng and S.O. City, “Study on the Rural Tourism Development in Bazhong City Based on Local Culture--A Case Study of Sicheng Village of Enyang District”, *Journal of Leshan Normal University*, **156(3)**, 178-191 (2017).
- Peng, KL. And P.M.C. Lin, “Social entrepreneurs: innovating rural tourism through the activism of service science”, *International Journal of Contemporary Hospitality Management*, **28(6)**, 1225-1244 (2016).
- Prince, S., “Craft-art in the Danish countryside: reconciling a lifestyle, livelihood and artistic career through rural tourism”, *Journal of Tourism and Cultural Change*, **178(3)** (2017).
- Qin, Y., Y. Luo, Y. Zhao, and J. Zhang, “Research on relationship between tourism income and economic growth based on meta-analysis”, *Applied Mathematics and Nonlinear Sciences*, **1(3)**, 105-114 (2018).
- Siow, M.L, S. Ramachandran, A. Shuib, et al., “Adapting evidence-based intervention in rural tourism policies”, *Worldwide Hospitality & Tourism Themes*, **7(5)**, 473-485 (2015).

Received: December 15th 2017

Accepted: June 30th 2018

Recommended by Guest Editor

Juan Luis García Guirao

