Research on Personalized Tourist Attraction Recommendation based on Tag and Collaborative Filtering

Yanqing Cui, Chuanlin Huang, Yanping Wang

School of Information and Business Management, Dalian Neusoft University of Information, Dalian 116023
E-mail: cuiyanqing@neusoft.edu.cn,huangchuanlin@neusoft.edu.cn,wangyanping@neusoft.edu.cn

Abstract: Tourists face a large number of tourist attractions, and spend a considerable amount of time and energy to select satisfactory tourist attractions. The application of personalized recommendation technology is an effective way to solve this problem. On the one hand, users' consumption frequency in tourism is much lower than other commodities such as music and movies; on the other hand, the increasing number of tourist attractions has led to the problem of sparse scoring data in personalized recommendations of tourist attractions. The traditional collaborative filtering algorithm is not satisfactory in the recommendation of tourist attractions. This paper builds a tourist attraction tag system, which links tourists and tourist attractions through the attractions tag from four aspects: location, location type, travel time, and travel method. By calculating the relationship between tourists and attractions tags, tourist attractions and attractions tags, a user interest model is constructed. Then, according to the user interest model, the interest degree of the new attraction to be recommended is predicted, and finally the tourist attraction recommendation set is generated.

Key Words: Tag , Collaborative filtering algorithm, Tourist attractions, Personalized recommendation

1 INTRODUCTION

With the development of the national economy, tourism has become a sunrise industry, and various types of tourism e-commerce platforms have emerged. Tourists encounter a “information overload” problem while obtaining massive amounts of information through the network, and it takes a lot of time to select. This not only increases the user's time cost, but also reduces the user experience. According to users' preferences, habits, personalized needs and the characteristics of goods, users' preferences for goods can be predicted. Online travel websites can recommend the most suitable products for users, help users make decisions quickly and improve users' satisfaction. Recommendation system, as an effective tool to solve information overload and meet users' personalized needs, has been successfully applied in Amazon and other e-commerce platforms. At present, recommendation systems are commonly used in the following categories: 1) recommendation based on collaborative filtering; 2) recommendation based on content; 3) mixed recommendation. Collaborative filtering recommendation technology is the most popular and widely used recommendation technology in current recommendation algorithms. The main idea is to divide users into different user groups according to their similar interests and hobbies, and recommend a particular user according to the preference habits of other users in the user group to which he belongs. Traditional collaborative filtering recommendation algorithms have achieved great success in e-commerce, video websites, news reading and other fields, but there are still many challenges when they are applied to the recommendation of tourist attractions. The characteristics of tourism consumption are low consumption frequency, insufficient user rating data and data sparsity.

In Web 2.0, tag is an important information resource. It is first marked by the public and widely recognized by users, and the tagged tags are classified, such as tourist attractions can be classified as: domestic tour, group tour, self-parade, etc. In this way, a tag-based information classification method is formed. Collaborative filtering algorithm reflects users' interest preference and resource similarity by user rating, ignores the characteristics of users and resources, and significantly reduces the quality of recommendation for sparse data and new resources. The Attractions tag can indicate the interest of the tourism and describe the characteristics of the tourist attraction. Introducing tag data into personalized recommendation of tourist attractions, combined with collaborative filtering recommendation algorithm, can effectively alleviate the problem of low recommendation performance of tourist attractions caused by the sparsity of scoring matrix, and improve recommendation diversity.

2 RECOMMENDER SYSTEM

The Recommendation System (RS) is a system tool that can automatically search, analyze, and calculate information and content of interest according to the user behavior. It obtains users' interests and preferences for resources by explicit or implicit ways, and recommends resources of interest to users. Applications for recommendation systems include e-commerce, social networking, movies, music, personalized email and advertising. Recommendation system generally consists of three modules, as shown in figure 1 below.

(1) User behavior data collection module. This module obtains user history behavior record from user log, such as shopping record, browsing record, score record and so on. This module is responsible for processing unstructured user behavior data into structured data, and it is the primary source of data for recommendation systems.

(2) User interest model building module. This module is responsible for extracting user characteristics, analyzing user data through machine learning knowledge, and creating a model for users. In the field of recommendation, different application backgrounds need to establish appropriate recommendation models.

This work is supported by Liaoning Doctoral research start-up fund guidance program project, project number: 20170520320

(3) Recommendation generation module. This module is the most important module in the whole system, is the core recommendation algorithm in the system, to a large extent determines the merits and demerits of the whole system. The recommendation module will select the appropriate recommendation algorithm to predict the score and generate recommendations.

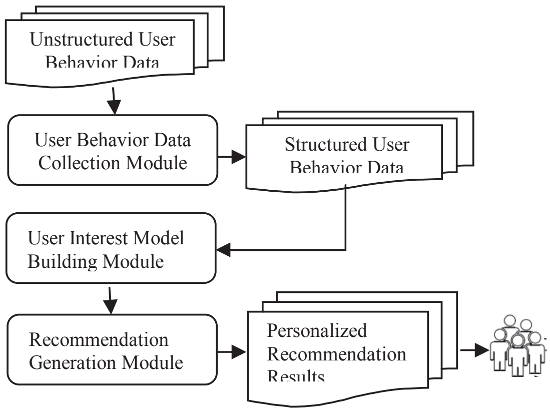

Fig 1.Personalized recommendation system model

3 PERSONALIZED RECOMMENDATION ALGORITHM

3.1 Collaborative Filtering Recommendation Algorithm

Collaborative filtering recommendation algorithm is the earliest and well-known recommendation algorithm. The algorithm discovers users' preferences by mining users' historical behavior data, divides users into groups based on different preferences and recommends similar resources. Collaborative filtering recommendation algorithms are divided into two categories: user-based collaborative filtering and item-based collaborative filtering.

(1) User-based collaboratIve filtering

User-based collaborative filtering algorithm discovers user's preferences for goods or content (such as purchase, collection, content review or sharing) through user's historical behavior data, and measures and scores these preferences. The relationship between users is calculated according to their attitudes and preferences for the same goods or content, and Resource Recommendation is made among users with the same preferences. In short, if both user A and user B have purchased items x,y and z, then both user A and user B belong to the same category of users. Therefore, user B can also be recommended to buy item w purchased by user A.

(2)Item-based collaborative filtering

Similar to the user-based collaborative filtering algorithm, the item-based collaborative filtering algorithm obtains the relationship between items by calculating the scores of different users for different items. Users are recommended similar items based on the relationship between items, and the score here represents the user's attitude and preference towards products. In short, if user A buys both item a and item b, then a and b are highly correlated. When user B also buys item a, it can be inferred that he also needs to buy item b.

3.2 Tag-Based Recommendation Theory

A tag is a non-hierarchical structure used to describe information and can be used to describe the semantics of an item. Tourists can use the attraction tag to describe their views on tourist attractions, so the attraction tag is the link between tourists and tourist attractions, and is also an important data source for responding to tourists' interests. Each tourist can mark multiple tourist attractions at the same time, and the set of attractions tag marked by each tourist can potentially show the interest and hobby of the tourist. An attractions tag can mark multiple tourist attractions at the same time, and the tag can reflect the common features of these attractions .An attractions tag can mark multiple tourist attractions, which can reflect the common characteristics of these tourist attractions. An attractions tag is used by multiple tourists, indicating that these tourists are interested in the same type of tourist attractions and have the same preferences.

A complete tag system usually consists of three elements: user, resource, and tag, as shown in Figure 2. In this paper, it is shown as the relationship among tourists, tourist attractions and attractions tag. Different from the collaborative filtering recommendation algorithm based on tourists' two-dimensional rating matrix for tourist attractions, the personalized recommendation algorithm based on tags relies on the three-dimensional relationship between tourists, tags and resources. It is necessary to decompose the three-dimensional matrix into two two-dimensional matrices, namely, the user-tag matrix and tag-resource matrix, as shown in Figure 3 and 4, and complete the recommendation process by replacing the tourist rating matrix with the importance data of tag.

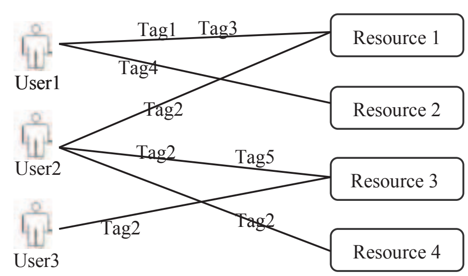

Fig 2. User-Tag-Resource Model

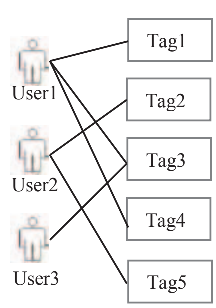

Fig 3. User-Tag Model

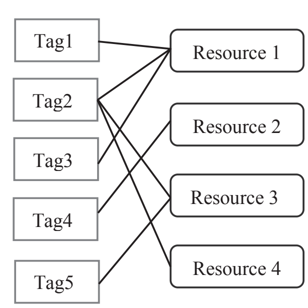

Fig 4. Tag-Resource Model

4 PERSONALIZED RECOMMENDATION MODEL OF TOURIST ATTRACTIONS BASED ON TAG AND COLLABORATIVE FILTERING

4.1 Tag System of Tourist Attractions

Many personalized recommendation websites provide users with social tag recommendation function, users can tag and classify resources based on custom tags. Tags are usually simple words or words, and the establishment of general tag system can be divided into two kinds. One method is to use social tag, which has the advantages of diversity and can reflect the user's personalized information, but it is prone to tagging errors and non-semantics issues; the other method is to establish a tag system by experts, users use the established tag system for tagging. In this paper, expert interview method is used to establish the tag system. Through general investigation and expert interviews, it is found that the evaluation of attractions by tourist enthusiasts is mostly related to the location, type, time and mode of travel of tourist attractions. Therefore, this paper constructs a tourist attraction tagging system consisting of 4 categories and 12 tags, as shown in figure 5 below.

According to their preference, tourists will tag the tourist attractions with personalized tag. For example, one tourist may tag Hangzhou West Lake with "natural landscape" and "free travel", while another may tag it with "package tour" and" vacation tour ". To some extent, the tag reflects tourists' travel preferences, and it is also convenient to classify tourist attractions.

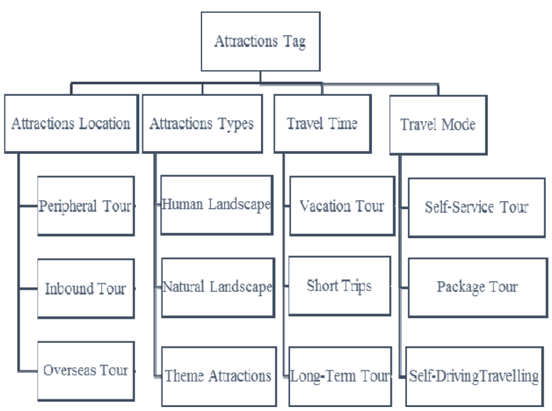

Fig 5. Tag System of Tourist Attraction

4.2 Personalized Recommendation Model

Different from the traditional collaborative filtering recommendation algorithm, the tag-based personalized recommendation of tourist attractions integrates the tag of tourist attractions into the personalized recommendation model, and uses the tag of tourist attractions as the intermediary to mine the interest of tourists from both aspects of tourists and tourist attractions. In order to get the TOP-N recommendation for tourists, it is necessary to calculate the preference degree and similarity degree of tourists for attractions tag. As shown in Figure 6.

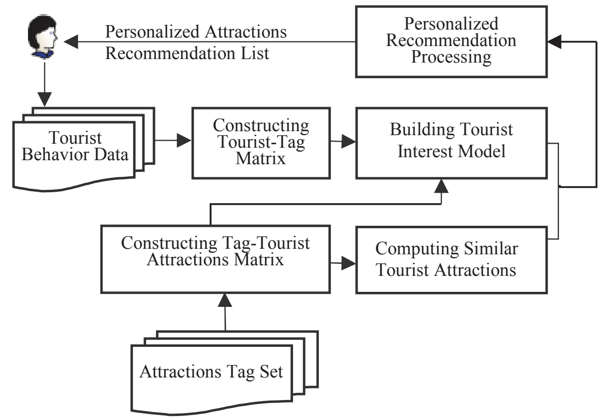

Fig 6. Personalized Recommendation Model of Tourist Attractions Based on Tag and Collaborative Filtering

Users tag tourist attractions with attractions tag, forming the behavior data of tourists, which constitutes the online data part of personalized recommendation system. The tag set of tourist attractions is formulated by experts in advance, which constitutes the off-line data part of the personalized recommendation system.

When users visit the website, the website is calculated according to personalized recommendation to recommend tourist attractions that conform to tourists' travel habits. The specific process is as follows:

(1)Constructing the model of tourists' interest. According to the historical behavior data of tourists, the tourist-attractions tag matrix and the tag- tourist attractions matrix are constructed to explore the interest model of tourists for tourist attractions.

(2) Computing the similarity of tourist attractions. Each tourist attractions may be marked with different tags by different tourists. The more tags are marked on tourist attractions, the more characteristics of tourist attractions can be reflected.

(3) form a top-n recommendation list. According to the interest model of tourists and similarity of tourist attractions, the degree of tourists' interest in tourist attractions is predicted, and a set of recommendation results is formed.

5 PERSONALIZED RECOMMENDATION ALGORITHM FOR TOURIST ATTRACTIONS BASED ON TAG AND COLLABORATIVE FILTERING

5.1 Constructing the Model of Tourists' Interest in Tourist Attractions

The tourist attractions tags are used as the characteristics of the tourist preference model, and the multi-dimensional relationship among tourists, attractions tags and tourist attractions is used to calculate the tourist preference model for tourist attractions.

Define tourist set $U=\{u_1, u_2, \cdots, u_i, \cdots, u_m\}$, where m is the total number of tourists, $i=1, 2, \cdots, m$; The set of all tourist attractions is $S=\{s_1, s_2\cdots, s_j, \cdots, s_n\}$, Where n is the total number of tourist attractions, , $j=1, 2, \cdots, n$; The set of all attractions tags is $T=\{t_1, t_2, \cdots, t_p, \cdots, t_k\}$, Where k is the total number of tags, $p = 1,2... , k$.

(1)Construct the Tourist- Attractions Tag Matrix

For different tourists, the more frequently they use a attractions tag l, the more they prefer the tourist attractions marked by the tag. TF-IDF method was used to calculate the importance parameters of attractions tags to tourists. TF-IDF can not only express the ability of classification of attractions tag, but also reflect the user's preference for tourist attractions to a certain extent. Tourist-attractions tag matrix is constructed by using historical behavior data of tourists, which reflects the possibility of tourists choosing the tag or the importance of the tag to tourists, and is represented by matrix P (as formula (1)), in which w_{ij} represents the importance of attractions tag t_j to the user i. as formula (1)

$$P = \begin{array}{c} \\ u_1 \\ \vdots \\ u_m \end{array} \begin{array}{c} \begin{array}{cccccc} t_1 & t_2 & \cdots & t_p & \cdots & t_k \end{array} \\ \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1p} & \cdots & w_{1k} \\ & \cdots & w_{ij} & & \cdots & \cdots \\ w_{m1} & w_{m2} & \cdots & w_{mp} & \cdots & w_{mk} \end{bmatrix} \end{array} \quad (1)$$

The value of w is calculated by the TFIDF method. Using TF-IDF to calculate the importance of tag to users and resources, it can reduce the weight of popular tags and resources and improve the novelty and personality of recommendation results The calculation method is shown in formula (2).

$$w_{ut_j}=\frac{n_{ut_j}}{n_{ut}}\log\frac{m}{n_{t_ju}} \quad (2)$$

In the formula, n_{t_ju} represents the number of users who use tag t_j , m represents the total number of users, n_{ut_j}represents the number of tags that user u uses tag t_j , n_{ut} represents the number of tags that user u uses, $\frac{n_{ut_j}}{n_{ut}}$ represents the frequency that user u uses tag t_j , and log $\log\frac{m}{n_{t_ju}}$represents the importance of tag among all tags of user.

(2) Construct Attractions Tag- Tourist Attractions Matrix

For a tourist attraction, it can be tagged by different users at the same time, and the tag can reflect the characteristics of the tourist attractions. If a tourist attraction tag is more important to a tourist attraction, the tag can reflect the characteristics of the tourist attractions. The attractions tag-tourist attractions matrix is represented by matrix Q(as formula (3)), where r_{ij} represents the importance of the attractions tag t_i to the tourist attractions j , or can be understood as the degree to which the tag describes the tourist attractions.

$$Q= \begin{array}{c} \\ t_1 \\ \vdots \\ t_k \end{array} \begin{array}{c} \begin{array}{cccccc} s_1 & s_2 & \cdots & s_q & \cdots & s_n \end{array} \\ \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1q} & \cdots & r_{1n} \\ & \cdots & r_{ij} & & \cdots & \cdots \\ r_{k1} & r_{k2} & \cdots & r_{kq} & \cdots & r_{kn} \end{bmatrix} \end{array} \quad (3)$$

The TF-IDF method is used to calculate the importance of attractions tag to tourist attractions, as formula (4).

$$r_{st_j}=\frac{n_{st_j}}{n_{st}}\log\frac{n}{n_{t_js}} \quad (4)$$

n_{t_js} represents the number of tourist attractions tagged by t_j, n represents the total number of tourist attractions, n_{st} represents the total number of tags marked on tourist attractions s, n_{st_j}represents the number of tourist attractions tagged by t_j , $\frac{n_{st_j}}{n_{st}}$ represents the frequency of tourist attractions tagged by t_j , and $\log\frac{n}{n_{t_js}}$ represents the importance of tag t_j among all tags of tourist attraction s.

(3)Construct Tourists' Interest Model of Tourist Attractions

The interest model of tourist I for tourist attractions j can be expressed by the following formula(5):

$$I(u_i, s_j)= \sum_{k=1}^{k} w_{ik} * r_{kj} \quad (5)$$

5.2 Calculate the similarity of tourist attractions

The similarity of tourist attractions indicates the degree of similarity between two different tourist attractions. The traditional collaborative filtering algorithm is based on the number or score of resources shared by different users, which ignores the characteristics of resources themselves. In view of this, it is also beneficial for users to discover new resources by taking the historical behavior of user marking resources as resource characteristic information and using it for similarity calculation. The similarity of resources can be calculated by cosine, modified cosine similarity or Pearson similarity. It is found that compared with e-commerce user rating data, tourist rating data is sparse, which reduces the accuracy of similarity calculation. Tourist enthusiasts and e-commerce consumers are different. They seldom repeat their consumption and often do not visit the same tourist attractions again in a short time. Recommendation diversity is particularly important in the recommendation of tourist attractions. In this paper, the importance of attractions tag is used to calculate the similarity between tourist attractions, instead of the traditional method of calculating the similarity by users' rating of resources. It's also good for users to discover new resources. It mainly based on the importance of tourist attractions i and tourist attractions j to a certain attractions tag, the higher the importance, the more similar the tourist attractions i and j. In this paper, the cosine similarity formula is used to calculate the similarity of tourist attractions, as shown in formula (6) .

$$sim(s_i, s_j) =\cos(\vec{s_i},\vec{s_j})=\frac{\sum_{t\in T} a_{t,i}\cdot a_{t,j}}{\sqrt{\sum_{t\in T} a^2_{t,i}}\sqrt{\sum_{t\in T} a^2_{t,j}}} \quad (6)$$

In the formula (6) , T is a set of tags for all tourist attractions. It is known that $T=\{t_1, t_2, \cdots, t_p, \cdots, t_k\}$, $a_{t,j}$ are the importance of tag i to tourist attractions j, and sim(s_i, s_j) represents the similarity between the tourist attraction i and j.

5.3 Personalized Tourist Attractions Recommendation Result Set

According to tourist interest and similarity of tourist attractions, the preference degree R($\mathbf{u}, s_j$) of tourist u for tourist attractions s_j is calculated, and the recommendation result set is formed, and Top-N attractions are recommended to tourists. The calculation formula is as formula (7).

$$R(\mathbf{u}, s_j) = \sum_{i=1}^{m} I(u, s_i) * sim(s_i, s_j) \quad (7)$$

6 APPLICATION OF PERSONALIZED RECOMMENDATION ALGORITHM

This paper selects 100 pieces of user data of a tourism website for experiment. The data includes tourist id, tourist attractions name, tourist attractions tag, tourist attractions rating and other information. According to the previous description of the algorithm, it is necessary to construct two matrices of tourists-tags and tags-attractions. The values of the matrices express the preferences of tourists for a certain tag and the description degree of tags to attractions, respectively. The letter U is used to represent tourists and TA is used to represent tourist attractions. The data were divided into two groups, one group of training data and the other group of experimental data. According to the training data, the tourists' interest preference matrix is calculated, and the similarity between the tourist attractions in the experimental data and the scored data is predicted, and the recommendation matrix is calculated. TFIDF algorithm is used to calculate and process the experimental data. This paper uses Java language to program, and obtains the corresponding TF-IDF tag importance matrix. Then, according to the previous tag-based model of tourists' interest in tourist attractions, some of the results of table 1 are obtained. The matrix records users' interests, and can reflect users' preferences for tourist attractions.

Table1. Tourists' Interest Preference Matrix

	TA 1	TA 2	TA 3	TA 4	TA 5	TA 6	TA 7
U1	0.0139	0.0107	0.0040	0.0039	0.0107	0.0048	0.0073
U2	0.0028	0.0039	0.0028	0.0031	0.0039	0.0205	0.0028
U3	0.0139	0.0107	0.0040	0.0039	0.0107	0.0048	0.0073
U4	0.0208	0.0139	0.0208	0.0030	0.0139	0.0043	0.0270
U5	0.0012	0.0012	0.0012	0.0129	0.0012	0.0030	0.0012

For attractions where visitors have not been rated or tagged with attractions, the results of the recommendations are formed by calculating the similarities between the attractions.The similarity calculation results are shown in Table 2, and the recommendation matrix is shown in Table 3. Through calculation, the recommendation results of tourists 1 to tourists 4 are obtained. According to the top-n recommendation principle, several tourist attractions with the highest predicted score for each tourist are recommended.

Table2. Similarity Matrix of Tourist Attractions

	TA 1	TA 2	TA 3	TA 4	TA 5	TA 6	TA 7
TA 1	1.000	0.408	0.5000	0.4082	0.4082	0.3536	0.6325
TA 2	0.408	1.000	0.0000	0.3333	1.0000	0.5774	0.5164
TA 3	0.5000	0.0000	1.0000	0.4082	0.0000	0.0000	0.6325
TA 4	0.4082	0.3333	0.4082	1.0000	0.3333	0.0000	0.2582
TA 5	0.4082	1.0000	0.0000	0.3333	1.0000	0.5774	0.5164
TA 6	0.3536	0.5774	0.0000	0.0000	0.5774	1.0000	0.4472
TA 7	0.6325	0.5164	0.6325	0.2582	0.5164	0.4472	1.0000

Table3. Recommended Results

	TA 1	TA 2	TA 3	TA 4	TA 5	TA 6	TA 7
U1	0.0325	0.0349	0.0172	0.0202	0.0349	0.0253	0.0328
U2	0.0177	0.0233	0.0072	0.0087	0.0233	0.0272	0.0203
U3	0.0325	0.0349	0.0172	0.0202	0.0349	0.0253	0.0328
U4	0.0624	0.0537	0.0495	0.0362	0.0537	0.0398	0.0704
U5	0.0099	0.0095	0.0078	0.0150	0.0095	0.0053	0.0086

7 CONCLUSION

The expansion of Internet data scale promotes the problem of information overload, and users hope to obtain fast and accurate information when visiting websites. This paper extracts four categories of tourist attractions tags, namely, attractions location, attractions type, travel time and travel mode, describes tourists' interests and preferences with attractions tag, and recommends personalized tourist attractions to tourists using tag-based collaborative filtering algorithm.

REFERENCES

[1] Wang Xiaoyun, Zhao Jing, Xu Zuoning, Research on User Interest Mining and Personalized Recommendation Based on Socialized Annotation. Journal of Modern Information, 2018, 38(07):67-73+80.

[2] HE Ming, YAO Kai-sheng, YANG Peng, ZHANG Jiu-ling, Collaborative Filtering Personalized Recommendation Based on Similarity of Tag Information Feature, Computer Science, 2018, 45(S1):415-422.

[3] Tang Xiaobo, Personalized Recommendation Based on Ontology and Tags, Information Studies:Theory & Application, 2016, 39(12):114-119.

[4] Liu Rujuan, Research on Personalized Recommendation Method Based on Tag Clustering and User Mode, Journal of Modern Information, 2016, 36(06):74-78+99.

[5] Wang Mengtian, Wei Jingjing, Liao Xiangwen, Lin Jinxian, Chen Guolong, Personalized Recommendation Algorithm Fusing Comment Tag, Journal of Frontiers of Computer Science and Technology, 2016, 10(10):1429-1438.

[6] Dong Yue-hua ,Liang Xue-lei, Collaborative Filtering Recommendation Algorithm Based on Tag Importance, Science Technology and Engineering, 2018,18(14):172-178.