Journal of Computers Vol. 32 No. 2, 2021, pp. 197-212
doi:10.3966/199115992021043202017

User Portrait-based Hybrid Recommendation Method of Web Services

Chao Ma*[1,2], Zhihao Wang[2], Pengyu Cheng[1], Xiaoqi Liu[1], Haining Yu[3], Hongguo Zhang[1]

[1] School of Software and Microelectronics, Harbin University of Science and Technology, Harbin, Heilongjiang, 150040, China

[2] School of Economics and Management, Harbin University of Science and Technology, Harbin, Heilongjiang, 150040, China

[3] Research Center of Computer Network and Information Security Technology, Harbin Institute of Technology, Harbin, Heilongjiang, 150001, China
{machao8396, wangzhihao_hust, 18714585696, xiaoqi102, zhg07}@163.com, yuhaining@hit.edu.cn

Received 22 June 2020; Revised 2 November 2020; Accepted 3 December 2020

Abstract. As an effective means to solve information overload, service recommendation has become one of the main research directions of the scholars belonging to service computing domain. However, there are still some problems in traditional service recommendation methods, such as low accuracy of recommendation result and cold start for new users. For this reason, this paper proposes a hybrid recommendation method of web services based on user portrait. Firstly, this method establishes the user portrait model by considering the user's natural attributes and interest attributes; Secondly, this method calculates the user portrait similarity on the basis of user portrait model and the user rating similarity on the basis of user rating matrix simultaneously, then combines social trust degree together by considering the trust relationship between users in user's social networks; finally, the user portrait model, user similarity and social trust degree are integrated into the hybrid recommendation method of web services, so as to obtained the more accurate recommendation result for users. The experimental results show that the proposed method improves the accuracy of the recommendation result, and effectively alleviates the cold start problem.

Keywords: service recommendation, user portrait model, user similarity, social trust degree

1 Introduction

With the rapid development of the Internet and IT technology, a lot of new service information, service applications and service resources will appear on the Internet every day, so information overload has become a serious problem. Specifically, when users need to select a web service from a mass of candidate web services with the similar functions but the different performance to meet their own needs, it is difficult for users to quickly select the most suitable service from multiple services. As the effective method for solving information overload, service recommendation method recommends services for users by analyzing their historical behavior data. At present, the service recommendation methods mainly include the following categories: content-based service recommendation [1], service recommendation based on collaborative filtering [2], service recommendation based on association rule [3], and hybrid service recommendation [4]. Among them, the service recommendation method based on collaborative filtering is one of the most widely used service recommendation methods. The traditional service recommendation method based on collaborative filtering mainly identify a group similar users for the target user by comparing the target user's rating behaviors with the rating behaviors of the other users, and this group similar users is regarded as the neighbor users, then on the basis of the ratings of the

* Corresponding Author

neighbor users on a set of services, predict the ratings of the target user on these corresponding services, at last the recommendation result are determined according to the comprehensive user ratings on these service. However, there are still some problems with the collaborative filtering recommendations, such as low accuracy of recommendation result and cold start for new users.

The traditional collaborative filtering algorithm relies on the user's ratings on various services, but in reality, the users usually cannot make a timely evaluation after using the services, even not willing to or forget to evaluate the used services, which make the historical evaluation data become sparse, then affects the accuracy of the recommendation result. In addition, for new users, they do not have any or have very few historical evaluation data, there is new user cold start problem. For solving these problems of collaborative filtering recommendation, scholars constantly conduct some research. Polatidis et al proposed a multi-level collaborative filtering recommendation method, which divides user similarity into different levels, and adds constraints to each level. This method improves the accuracy of similarity measure and thus improves the accuracy of the recommendations [5]. Rong Huigui et al proposed a collaborative filtering recommendation algorithm based on user similarity, which redefined the composition and calculation method of user similarity in social networks. This method improves the accuracy of friend recommendation in social networks [6]. Wang et al proposed a collaborative filtering recommendation method of using social information, which comprehensively considers the user similarity relationships and the social trust relationships, enhances the computation of the user's neighborhood and improves the accuracy of the recommendations [7]. Bedi et al incorporated the concept of dynamic trust into the collaborative filtering recommendation, and selected the neighborhood based on biological metaphor of ant colonies. At the same time, new users can benefit from pheromone updating strategy. This method improves the accuracy of recommendation and alleviates the new user cold start problem [8]. The above mentioned literatures mainly increase the accuracy of collaborative filtering recommendations by improve the calculating method of user similarity, or by improving the selecting method of neighborhood. Although these literatures achieved some results, the lack of in-depth analysis on user feature information results in that the recommendation results is far from satisfactory.

Further, in order to get the more accurate recommendation result and better alleviates the cold start problem, more and more scholars have introduced the concept of user portrait into the service recommendation method, and achieved some results. Saia et al identify and remove the semantically incoherent items from the previously evaluated items, so that the user portrait was closer to the user's real preferences, and the accuracy of the recommendations was improved [9]. Wang Dong builds a user portrait model based on user personal information and rating information, analyzes the dimensions of the user portrait in the book purchase scene, and incorporates the user portrait on the basis of content-based recommendation to achieve the accurate recommendation [10]. Liu Yong et al propose a recommendation method based on dynamic user portrait, in which the user's historical data were analyzed dynamically to predict the changes in user preferences, and finally improves the recommendation accuracy [11]. The above mentioned literatures establishes the user portrait based on the various user feature information, which is mainly an improvement on the content-based recommendation method. However, this paper mainly focus on how to solve the problems of traditional collaborative filtering recommendation: low accuracy and new user cold start. Therefore, we needs further improvement on the construction method of the user portrait model, and then incorporate it into the collaborative filtering recommendation.

From the above, when the collaborative filtering-based recommender system is executing, only considering the user's rating similarity, there are still the problems: (1) the rating data is sparse, and at the same time, the timeliness of rating data would become decrease over time, which would affect the accuracy of the recommendation result, and (2) for new users, they don't have any or have very little rating data, which would make new users cannot obtained the appropriate recommended service. These are the key research problem that we want to solve in this paper. Therefore, this paper introduces user portrait model into the traditional collaborative filtering algorithm. When the target user is a new user, the similarity between the new user and the old user is the user portrait similarity that is calculated according to the established user portrait model, some old users with a higher portrait similarity are selected, and then the recommendation results related to these old users are synthesized and used as the target user's recommendation result. When the target user is an old user, the user similarity consist of the user portrait similarity on the basis of user portrait model and the user rating similarity on the basis of user rating matrix, and then, based on the information such as the interaction and the interaction path between users, direct trust degree and indirect trust degree between users are calculated, and the direct and indirect trust degree are integrated to obtain the social trust degree between users. N users with higher similarity and social trust degree are regarded as the trusted neighbor users of the target user, so as to obtain recommendation result.

The technical contributions of the paper are summarized as follows:

(1) We cluster the tags in the tag library to obtain the representative K tag clusters, these tag clusters are used to reflect the interest attributes of different users.

(2) We establish the user portrait model, this model mainly includes the user's natural attributes and interest attributes. On the basis of the user portrait model, the user's natural attribute similarity and interest attribute similarity are calculated to obtain the user portrait similarity between users.

(3) Combining the user portrait similarity with the user's trust relationship in the social network and the user rating similarity considering timeliness, we propose a hybrid recommendation method of Web service based on user portrait.

The rest of the paper is organized as follows: Section 2 presents the method of constructing user portrait model; in Section 3 we introduce the hybrid recommendation method based on user portrait; Section 4 proves the feasibility of the proposed method by experimental results; Section 5 contains conclusions and future work.

2 User Portrait

User portraits are the virtual representatives established based on real users. This paper mainly builds user portrait models from two aspects: natural attributes and interest attributes. Among them, the natural attributes refer to users' basic information and the interest attributes are obtained based on users' interest tags. The specific methods for creating user portrait models are given below.

2.1 Label Clustering

Clustering is the process of dividing a collection of physical objects or abstract objects into multiple classes which composed of similar objects. For the multiple classes generated by clustering, their data characteristics of the objects belonging to the same class are similar. Commonly used clustering methods include: K-means algorithm, hierarchical method, model algorithm, density algorithm, grid algorithm, etc. Among them, the K-means algorithm is the most classic. This paper uses the optimized bipartite K-means clustering method proposed in reference [12] to cluster the tags in the tag library. In the calculation method of the distance between labels, using the similarity between the labels to replace the traditional Euclidean distance to cluster the labels.

Suppose there are l tags in the data, expressed as $T=\{t_1,t_2,\ldots,t_l\}$, and there are n services, expressed as $S=\{s_1,s_2,\ldots,s_n\}$. The similarity between tags should be related to the tagging service status and tag co-occurrence status. Label co-occurrence refers to a situation where two or more labels jointly mark a service. If the number of label co-occurrence is greater, it means that the correlation between the two tags is greater. The calculation of familiarity $V\_(i,j)$ between tags is given below, as shown in formula (1).

$$V_{i,j}=\frac{N(t_i,t_j)}{\Sigma_{i,j\in l}N(t_i,t_j)}\times lg\frac{n}{1+C(t_i)}. \tag{1}$$

In the formula (1), $V_{i,j}$ represents the familiarity between the tags t_i and t_j, that is, the degree of association between the two tags. $N(t_i,t_j)$ represents the number of times the tag t_i and tag t_j jointly mark the service; n is the total number of tags; $C(t_i)$ represents the number of other tags that have been tagged together with the tag t_i. It can be seen from the formula that when the frequency of co-occurrence labels between labels is greater and the frequency of co-occurrence with other tags is less, the familiarity between the tags is higher. According to the familiarity between the calculated labels, the relationship matrix T(l, l) between the labels is established and it's shown as follows.

$$T(1,1)=\begin{bmatrix} V_{1,1} & V_{1,2} & \cdots & V_{1,1} \\ V_{2,1} & V_{2,2} & \cdots & V_{2,1} \\ \cdots & \cdots & \cdots & \cdots \\ V_{1,1} & V_{1,2} & \cdots & V_{1,1} \end{bmatrix}. \qquad \textbf{(2)}$$

Each row in the matrix represents the familiarity between a label and all other labels, which can be abstracted as the relationship vector $T_i=(V_{i,1},V_{i,2},V_{i,3},\dots,V_{i,l})$ between the labels. According to the cosine similarity calculation formula, the similarity $sim(t_i,t_j)$ between the two labels is obtained. The calculation method is shown in formula (3).

$$sim(t_i,t_j)=\frac{\overline{T_I}\cdot\overline{T_J}}{\|\overline{T_I}\|^2\times\|\overline{T_J}\|^2}. \qquad \textbf{(3)}$$

The optimized binary K-means clustering method flow is as follows:

1. Set the number of clusters to K;
2. Initialize and put all the labels as a cluster into the cluster set;
3. Take out the cluster with the largest sse in the cluster set, $sse=\sum_{i=1}^{k}\Sigma_{x\in c}sim(c_i,x)^2$, in this expression k is the number of clusters, c_i is the center of the first cluster, and C is the set of all cluster centers. Choose a label from the cluster with the largest see, calculate the label with the lowest similarity to this label, and record it as C_1. Then calculate the point with the lowest similarity to label C_1, and record it as C_2.
4. By calculating the similarity of all labels to C_1 and C_2, put the labels into the clusters centroid with the highest similarity to form the two clusters T_1 and T_2. And then add these two clusters to the cluster set.
5. Repeat steps 2 to 4 until the cluster contains K clusters;
6. Taking the center of K clusters in the cluster set as the initial centroid, the average clustering is performed on all points to obtain the final K clusters which are expressed as $\{T_1, T_2, T_3, \dots T_K\}$.

2.2 User Portrait

User portraits were first proposed by Alan Cooper, the father of interaction design [13]. He said: “The user portrait is a specific representative of the target user.” That is, user portrait is the virtual representative established for real users and it is the target user model established on the basis of real information data. The dimensionality design of user portrait is to tagging the user and it is the basis of user modeling. It mainly includes three dimensions of natural attributes, interest attributes and ability attributes. The following figure shows the dimensionality of user portraits. As shown in Fig. 1.

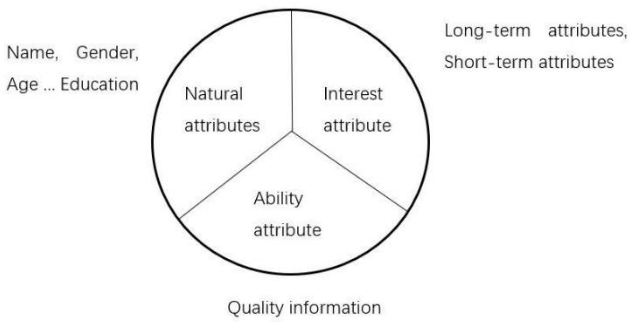

Fig. 1. User portrait dimension

Natural attributes: it refers to the attributes possessed by the users themselves, that is, the natural attribute is an attribute generated innately or acquired to represent the user’s basic information. Generally, it is the basic information of users when they register, including gender, age, address, education background, etc. The user characteristics represented by natural attributes are basically unchanged feature labels, but different features have different effects in different practical application scenarios, so each feature will occupy different weights in different environments.

Interest Attribute: it refers to the user's interest preference expressed in a specific environment. The characteristic information about the user's interest can be extracted by user's online purchase, evaluation and other behavior information. The interest attributes plays an important role in the user's portrait and the user's specific operation can reflect the user's characteristics best.

Ability attribute: it refers to the user's ability to reflect on the network and whether the user can give thoughtful, insightful and more influential evaluation information. For example: Stars and "V" users in Weibo have greater influence.

The user profile model can be established according to the three dimensions of the user profile's natural attributes, interest attributes, and ability attributes. This paper mainly studies the first two dimensions, modeling user portraits from natural attributes and interest attributes. The modeling process is as follows.

Define a two-dimensional tuple to represent the two dimensions of the user portrait: natural attributes and interest attributes. The definition of the tuple is as follows:

User = <Natural, Interest>

In this tuple, the "Natural" represents the user's natural attribute dimension and the "Interest" represents the user's interest attribute dimension. The natural attributes dimension is mainly the information filled in when the user registers, it includes age, gender, education, occupation, etc. It is usually easily collected, basically stable, and will not change in a short time, that is, the natural attributes doesn't need to be updated in a short time. The interest attributes dimension is summarized by the user's behaviors over a period of time, so the interest attributes are relatively complex, difficult to collect, and dynamically changing. The following describes the characteristic information in each dimension.

All characteristics of the user's natural attribute dimension:

- Name: the name filled in when the user registered.
- Age: the user's age.
- Gender: 0 represents male and 1 represents female.
- Address: user's area.
- Occupation: the user's job.
- Education: user's education, including junior college, undergraduate, master, doctor, etc.
- Personal introduction: the user's self-introduction can include basic information such as personality, hobbies and so on.

The user's interest attributes dimension mainly using the user's behavior records to summarize the user preferences. This user portrait model mainly builds interest attributes based on the interest labels clusters obtained by clustering u_i, $u_i = \{T_1, T_2, T_3, \ldots T_K\}$. There are two cases: (1) If the user is an old user, this paper mainly determines the interest attribute based on the label situation used by the user. According to the above introduction of tag clustering, there are K interest tag families in total, and the user's interest is determined according to the distribution of tags used by the user in each interest tag family; (2) If the user is a new user, the user's interest attributes are mainly determined according to the interest tag cluster that the new user chooses when using the system.

The following gives a schematic diagram of the user portrait model, as shown in Fig. 2.

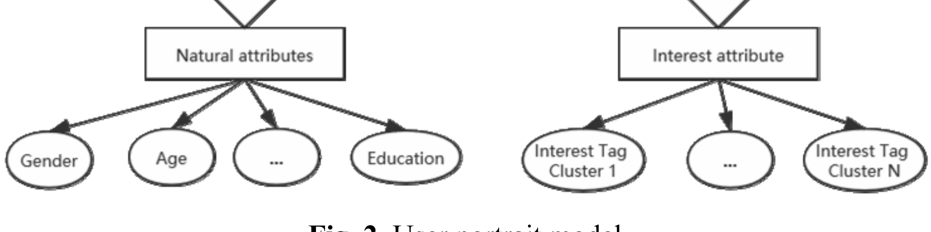

Fig. 2. User portrait model

The user portrait model is designed in a layered manner, which can easy to add or delete the situation in each layer. For example, there are some features in the user's natural attributes that don't have much reference value, therefor when calculating the similarity between users, some features don't need to be considered.

3 Hybrid Recommendation Method Based on User Portrait

3.1 User Portrait Similarity

The similarity based on the user portrait refers to calculating the similarity between the new user and the old user based on the established user portrait model. The user portrait similarity on the basis of user portrait model mainly including both the natural attribute similarity and the interest attribute similarity. Specific calculation methods are given below.

According to the user portrait model, natural attributes include information such as name, age, gender, etc., which can be expressed as a vector Unatural = $\{n_1, n_2, \ldots, n_n\}$, in this vector n represents the number of features of the natural attributes of the user portrait. But due to some information in the natural attributes has no effect on the calculation of the similarity between users, so this paper selects three characteristics of gender, age and education as the feature information for calculating the similarity of natural attributes between users. The formula (4) for calculating the similarity of users based on natural attributes is shown in below.

$$sim\_n(u_i, u_j) = a \cdot sim(u_i, u_j, n_1) + b \cdot sim(u_i, u_j, n_2) + c \cdot sim(u_i, u_j, n_3). \quad \textbf{(4)}$$

In this formula, $sim\_n(u_i, u_j)$ represents the similarity between the user u_i and the user u_j regarding the natural attributes. a represents the weight of the gender feature, $a = 0.15$. $sim(u_i, u_j, n_1)$ represents the gender similarity between the user u_i and the user u_j, that is, whether the gender between the users is the same, when the gender is the same, $sim(u_i, u_j, n_1)$ takes the value 1; otherwise, it takes the value 0. b represents the weight of age feature, $b = 0.35$. $sim(u_i, u_j, n_2)$ represents the age similarity of user u_i and user u_j, when |user_uiage-user_ujage| > 6, $sim(u_i, u_j, n_2)$ takes the value 0; when 3 < |user_uiage-user_ujage| ≤ 6, $sim(u_i, u_j, n_2)$ takes the value 0.4; when 1 < | user_uiage-user_ujage| ≤ 3, $sim(u_i, u_j, n_2)$ takes the value 0.8; when | user_uiage-user_ujage| ≤ 1, $sim(u_i, u_j, n_2)$ takes the value 1. c represents the weight of the academic attribute, $c = 0.5$. $sim(u_i, u_j, n_3)$ represents the similarity of the academic qualifications of the user u and the user v. When the educational levels between the users are the same, $sim(u_i, u_j, n_3)$ takes the value 1; when the academic level between two users is different, the degree levels of users differ by one level, the value of $sim(u_i, u_j, n_3)$ will be correspondingly reduced by 0.2. For example, user A is an undergraduate and user B is a graduate student, then $sim(u_i, u_j, n_3)$ takes a value of 0.8; user A is a junior college and user B is a graduate student, then $sim(u_i, u_j, n_3)$ takes a value of 0.6.

According to label clustering, the labels used by old users are aggregated into K different types of interest label clusters, which are expressed as $\{T_1, T_2, T_3, \ldots, T_K\}$, that is, T1 to TK represent K interest label clusters. The user's liking for each interest tag cluster is different. In the interest attributes of the user portrait model, each user's interest tag cluster is sorted. New users reflect their interests by selecting interest tag clusters and old users reflect interest attributes based on the distribution of used tags in each interest tag cluster. When comparing whether the new user and the old user are similar in the interest dimension, in addition to calculating whether the selected interest label clusters are similar, it is also necessary to consider whether the two users' liking level of the interest label clusters is similar. Specifically, the similarity of the users' liking level for the interest label cluster is determined according to the frequency of the tags used by two users in the same interest tag cluster. The following gives an expression of the similarity degree L that the users' liking level to interest label clusters, as shown in formula (5).

$$L = e^{\Sigma - |n_i - n_j|}. \tag{5}$$

In the formula (5), u_i and n_j respectively represent the number of labels which user u_i and user n_j in a certain interest label cluster. Since the new user has not used tags for labeling, the number of new user's label is set as the half number of the total in each selected interest label cluster. When $n_i = n_j$, , L = 1, it means that the two users have the same number of labels in each interest label cluster and the two users' liking level for labels is the same.

According to the user's label vector and the user's similarity to the label's liking level, the interest similarity between users is obtained, as shown in formula (6).

$$sim\_i(u_i, u_j) = L \cdot \frac{T_i \cdot T_j}{\| T_i \| \times \| T_j \|}. \tag{6}$$

In the formula (6), T_i, T_j represent vectors of interest label clusters that the user is interested in.

The similarity of new users' portraits and old users' portraits mainly includes two aspects: natural attribute similarity and interest attribute similarity. According to the similarity between the two dimensions of natural attribute and interest attribute calculated above, the formula for the user's final user portrait similarity is shown in (7).

$$sim\_p(u_i, u_j) = \gamma sim\_n(u_i, u_j) + (1-\gamma) sim\_i(u_i, u_j). \tag{7}$$

In the formula (7), γ represents the weight of natural attribute similarity, and $(1-\gamma)$ represents the weight of interest attribute similarity.

3.2 User Rating Similarity

After using a service, users will give a corresponding evaluation according to their own experience. But for users, everyone has their own rating habits, some people are used to giving high scores, and some people are used to giving lower scores. Since the scoring criteria are different, only looking at the service's grading level can't compare the similarity between users. Therefore, it is impossible to compare the similarity between users only by looking at the level of the scores which are given by users. It is also necessary to process the original scores to obtain the user's objective scores of the service and then use the objective scores form a user rating matrix to obtain the user rating similarity which is on the basis of user rating matrix.

Suppose the set of users is U and $U = \{u_1, u_2, ..., u_m\}$, the set of all services is S and $S = \{s_1, s_2, ..., s_n\}$, using $r_{i,k}$ to represent user u_i's rating of service s_k. Since the data score level range in the generally recommendation method data set is on 0 to 5, $r_{i,k}$ is an integer from 0 to 5. If the user u_i doesn't rate the service s_k, then $r_{i,k} = 0$. In order to eliminate the user's personal rating habits, the user's objective rating $w_{i,k}$ is obtained according to formula (8), and its value is between 0 and 1.

$$w_{i,k} = \frac{r_{i,k}}{\sqrt{\sum_{i=1}^{S} {r_{i,k}}^2}}. \tag{8}$$

Using the formula (8), the objective rating value $w_{i,k}$ which eliminates the user's own rating habits can be obtained. By using this formula makes the ratings of each user under the same standard, and improves the accuracy of the rating.

The effectiveness of the user score for the services is related to the time of the user's rating. Ratings farther away from the current time do not reflect the user's preferences very well, and the user's interest will gradually decline over time. Therefore, when using the user's rating, the time-validity of the rating should be fully considered to make the data more accurate.

The closer the rating time is to the current time, the more the user's preference can be reflected, and the farther the rating time is from the current time, the lower the reference value of the score. Considering the time of rating service and the life cycle of user rating behavior, the service score timeliness function is obtained, as shown in formula (9).

$$w_{time} = \begin{cases} \exp\{-\ln 2 \times time / \max\} & time \le \max \\ \dfrac{1}{2} & time \ge \max \end{cases}. \quad \textbf{(9)}$$

In the formula (9), w_{time} represents the timeliness of user u_i scoring to the service s_k, time is a non-negative integer, let L be a fresh cycle of scoring time. When the time between user u_i's evaluation time to service s_k to the most recent evaluation time is 0-L, the value of time is 0. When the time between user u_i's evaluation time to service s_k to the most recent evaluation time is L-2L, the value of time is 1, and so on. When the evaluation time is the closer to the most recent evaluation date, the smaller the value of time. $\max$ is a threshold, if the larger the value of max, the slower the decline in user interest, and vice versa. When $time \ge \max$, the value of w_{time} will remain at 1/2. Therefore, the closer the time the service is evaluated, the higher the time score, and the longer the time the service is evaluated, the lower the time score.

This section improves on the shortcomings of traditional scoring similarity calculation methods, and comprehensively considered the personal scoring habits and scoring timeliness. Based on comprehensive consideration of personal scoring habits and timeliness of scoring, the total score of user u_i's preference for service s_k is obtained, the specific formula is shown in (10).

$$N_{i,k} = \lambda w_{i,k} + (1-\lambda) w_{time}(i,k).. \quad \textbf{(10)}$$

In the formula (10), λ is the weight occupied by $w_{i,k}$, and its value is between 0 and 1. The user scoring for the service can reflect the user's preference well and scoring for the time can reflect the change of the user's preference better. In general, taking λ = 0.5 means that the two are equally important.

According the above-mentioned scoring define a user-service scoring matrix which represents the comprehensive rating of the service s_k by the user u_i. The matrix is represented by formula (11).

$$N(m \times n) = \begin{bmatrix} N_{1,1} & N_{1,2} & \cdots & N_{1,n} \\ N_{2,1} & N_{2,2} & \cdots & N_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ N_{m,1} & N_{m,2} & \cdots & N_{m,n} \end{bmatrix}. \quad \textbf{(11)}$$

The similarity between users is calculated according to the above scoring matrix. In this paper, using the cosine to calculate the similarity between users. The calculation of similarity is shown in formula (12).

$$sim\_r(u_i, u_j) = \frac{\Sigma_{k \in s_{i,j}} N_{i,k} \times N_{j,k}}{\sqrt{\Sigma_{k \in s_{i,j}} N_{i,k}^2} \times \sqrt{\Sigma_{k \in s_{i,j}} N_{j,k}^2}}. \quad \textbf{(12)}$$

3.3 Calculation of Social Trust Degree Between Users

The trust relationship between users in social networks includes direct trust relationship and indirect trust relationship, as shown in Fig. 3. In Fig. 3, there is a direct trust relationship between user U1 and user U3, and an indirect trust relationship between user U1 and user U2.

Definition 1. Social relationship direct trust refers to a trust relationship in which the recommended user and the target user are friend relationships in a social network and can communicate directly. As shown in Fig. 3, there is a direct trust relationship between user U1 and user U3.

In social networks, the direct trust degree between users is related to the familiarity degree and core degree of social relationships. The calculation methods of the user's familiarity degree and the core degree in social networks are given below.

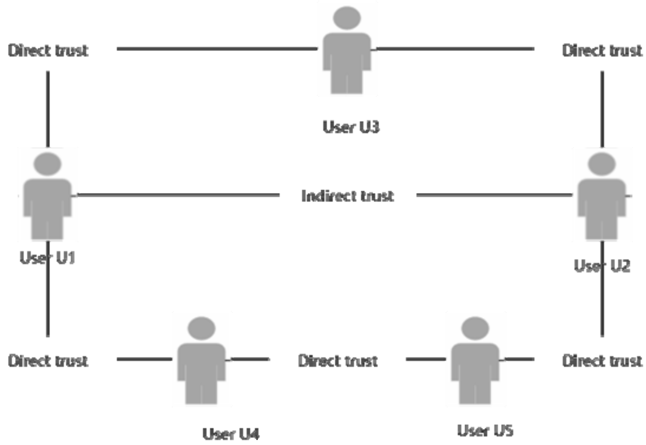

Fig. 3. Trust relationship in social network

The familiarity degree between users is usually related to the situation of interaction between users, the users are more familiar with friends who often contact and interaction. The calculation method of the familiarity degree between users is shown in formula (13):

$$F(u_i,u_j)=\frac{N_{i,j}-N_i\min}{N_i\max-N_i\min}. \quad \textbf{(13)}$$

In the formula (13), $F(u_i,u_j)$ is the familiarity degree of the social relationship between user u_i and user u_j, $N_{i,j}$ is the number of interactions between user u_i and user u_j, and $N_i\min$ is the minimum number of interactions between user u_i and other users, $N_i\max$ is the maximum number of times user u_i interacts with other users. Therefore, the more frequent the interactions between users, the more familiar they are.

The core degree of social relationship refers to the importance of a user in a social network. If a user is the core person in a social network, then it has a greater influence on other members. Therefore, the services recommended by core members are often believed by more people. The calculation of core degree is shown in formula (14).

$$C(u_j)=\begin{cases}1,|\,d_j=d_{\max}\\ lb(1+\dfrac{d_j}{d_{\max}}),|\,0<d_j<d_{\max}\\ 0,|\,d_j=0\end{cases}. \quad \textbf{(14)}$$

In the formula (14), $C(u_j)$ represents the core degree of the social relationship of the user u_j, d_j represents the number of friends that the user u_j has, and $d_{\max}$ represents the maximum number of friends of the user in the social relationship. The greater the number of friends, the higher the user's influence.

According to the above formula, social relationship familiarity degree $F(u_i,u_j)$ and social relationship core degree $C(u_j)$ can be calculated between the target user u_i and recommended user u_j. In order to get the recommended user u_j's direct trust degree for the target user u_i, using the formula (15) which is shown as follows.

$$Tru\_d(u_i,u_j)=\alpha F(u_i,u_j)+(1-\alpha)c(u_j). \quad \textbf{(15)}$$

In the formula (15), α is a weighting factor that reflects the familiarity and core degree in the social relationship, and the set value is different according to different industry services. It can be seen from the above that the direct trust degree of the social relationship is affected by the following two aspects. (1)

The interaction situation between target users and recommended users. In social networks, the more the number of interactions between users and the higher the frequency of interactions, the more intimate and familiar the users are, and then the lower the situation of that recommended user is the malicious user, the higher the direct trust degree. (2) The position for the recommended user in the social network. In the social network, some people belong to the core figures. As core figures, they receive more attention and have higher requirements for their own words and deeds, so they rarely make malicious comments.

Definition 2. Social relationship indirect trust refers that the recommended user and the target user are not a direct friend relationship in the social network, but establish contact through some joint friends. So the trust relationship between them is a kind of trust relationship established though passed on by several direct trust relationships.

As shown in Fig. 3, user U1 and user U2 have no direct contact and there is no direct trust relationship, but they can connect through user U3, or they can connect through user U4 and user U5. Therefore, there is an indirect trust relationship between user U1 and user U2.

As can be seen from Fig. 3, there can be multiple paths between the recommended user and the target user, the intermediate user can be one or more. Therefore, the user's indirect trust is related to the length of relationship path, number, the path weights and the direct trust degree of intermediate user. Let $Tru\_id(u_i, u_j)$ represent the indirect trust degree between the target user u_i and the recommended user u_j. The calculation method of the user's indirect trust degree is shown in formula (16).

$$Tru\_id(u_i, u_j) = \frac{\sum_{k=1}^{n} e^{-\lambda p_k} \prod_{l=1}^{p_k} DT_1(x, y)}{\sum_{i=1}^{n} e^{-\lambda p_k}}.. \quad \textbf{(16)}$$

In the formula (16), $e^{-\lambda p_k}$ represents the weight coefficient of the k-th relation path, and the value range is between 0 and 1. λ represents the attenuation coefficient of the relationship path length, p_k represents the length of the k-th relationship path. According to the trust maximum propagation distance calculated above obtaining $p_k \le d_{\max}$. Suppose there are n relationship paths between user u_i and user u_j, expressed as $p_{ij} = \{p_1, p_2, ..., p_n\}$, then the weight of each relationship path is $\{e^{-\lambda p_1}, e^{-\lambda pk_2}, ..., e^{-\lambda p_n}\}$. $DTl(x, y)$ represents the direct trust between users ux and uy in the k-th path and set D is a collection of all users connecting recommended users u_j and target users u_i in a path, users u_i and u_j are in the set D, that *ux*, $u_x, u_y \in D$.

Definition 3. Social trust degree is a comprehensive measure of trust relationship between users. It takes into account the direct trust degree and indirect trust degree between users. The social trust degree is abbreviated as $Tru(u_i, u_j)$, and its measurement method is shown in formula (17).

$$Tru(u_i, u_j) = \mu Tru\_d(u_i, u_j) + (1-\mu) Tru\_id(u_i, u_j). \quad (17)$$

In the formula (17), μ reflects the proportion of user direct trust and indirect trust, $\mu \in [0, 1]$. When $\mu = 1$, the social trust degree between users all comes from the direct trust relationship and there is no indirect trust; when $\mu = 0$, there is no direct connection between the two users, and the social trust degree all comes from the indirect trust relationship.

3.4 Hybrid Recommendation Algorithm Based on User Portrait

Firstly, combine the similarity based on the user portrait and the similarity based on the rating which are calculated above to obtain the similarity between users. The calculation method is shown in formula (18).

$$sim(u_i, u_j) = \beta sim(u_i, u_j) + (1-\beta) sim\_r(u_i, u_j). \quad \textbf{(18)}$$

In the formula (18), β reflects the proportion of user portrait similarity and rating similarity, $\beta \in [0, 1]$.

The existing service recommendation methods have problems such as low accuracy of recommendation result and cold start of users. The Web services trusted hybrid recommendation method based on user portrait proposed by this paper solves the above problems. When the target user is an old user, the similarity and trust between users are mainly calculated based on the information left by the user using the service, and the influence between users is calculated based on the similarity and trust, select N users with higher influence on the target user as the trusted neighbor users of the target user, so as to realize the recommendation. When the target user is a new user, since the new user has not used the service and has not conducted any behavior such as evaluation. So calculating user portrait - based similarity between the new user and the old user, selecting an old user which has a highest similarity and directly recommend the recommendation result of the old user to the target user. The specific process is as follows:

When the target user is an old user, the trusted neighbor users of the target user are selected based on the similarity $sim(u_i, u_j)$ and trust $Tru(u_i, u_j)$ between the users calculated above. The trusted neighbor users mainly select the users who have relatively high similarity and trust to the target user, calculates the comprehensive influence $affect(u_i, u_j)$ of the recommended user u_j on the target user u_i, and then finds the trusted neighbor users of the target user. The method of calculating the comprehensive affect value $affect(u_i, u_j)$ of the recommended user u_j on the target user u_i is given below, as shown in formula (19).

$$affect(u_i, u_j) = \frac{2sim(u_i, u_j) \times Tru(u_i, u_j)}{sim(u_i, u_j) \times Tru(u_i, u_j)}. \quad \textbf{(19)}$$

In the formula (19), $sim(u_i, u_j)$ is the similarity based on service score calculated using formula (17) and $Tru(u_i, u_j)$ is the social trust degree between users calculated using formula (16). When $sim(u_i, u_j) = 1$, $Tru(u_i, u_j) = 1$, the recommended user u_j 's influence on the target user u_i is 1, which is the largest. When $sim(u_i, u_j)$, $Tru(u_i, u_j) = 0$, the recommended user u_j 's influence on the target user u_i is 0, which is the lowest. When $sim(u_i, u_j) \in (0,1)$, $Tru(u_i, u_j) \in (0,1)$, the larger their value is, the higher the recommended user u_j 's comprehensive influence value on the target user u_i is.

After the above calculations, a group of N users with the highest influence on the target user is selected and regarded as a trusted neighbor user, denoted as G. According to trust neighbor users, the scoring prediction and service recommendation are realized. The calculation method of the predicted value $p_{i,k}$ of the user's service scoring is given below, as shown in formula (20).

$$p_{i,k} = \overline{\gamma_i} + \frac{\Sigma_{j \in G} affect(u_i, u_j)(r_{j,k} - \overline{\gamma_j})}{\Sigma_{j \in G} affect(u_i, u_j)}. \quad \textbf{(20)}$$

In the formula (20), $p_{i,k}$ represents the target user u_i 's predictive evaluation of the service s_k, $\gamma_{j,k}$ is the user u_j 's evaluation of the service s_k, $\overline{\gamma_i}$ and $\overline{\gamma_j}$ respectively are the average scores of users u_i and u_j for all the services used by their neighbor users.

For the service s_k, that the target user has not used, calculate the predicted score $p_{i,k}$ of the target user u_i for the service s_k. Using the top K services with higher predicted scores as user u_i 's service recommendation set, and the service recommendation ends.

When the target user is a new user, the recommendation result of the old user with a higher similarity to the new user is directly recommended to the new user as a service recommendation set.

The specific steps of the Web service hybrid recommendation method based on user portrait are as follows:

Input: target user u_i, user service rating R(m×n), tag set T, user friend relationship, time stamp

Output: Service recommendation set of target user u_i

1. Cluster the tags in the tag library into K tag clusters by clustering method, the clusters are express as $\{T_1, T_2, T_3, \ldots T_K\}$;

2. Establish user portrait models which include natural attributes and interest attributes. New users determine their natural attributes and interest attributes based on the basic information filled in when users register and the selected interest tags; old users determine their natural attributes and interest attributes based on the basic information filled in during registration and the scoring behavior information after using the service;
3. According to the user's gender, age and other characteristics, the natural attribute similarity $sim(u_i, u_j)$ between the new user and the old user is calculated by formula (4);
4. According to the interest label cluster selected by the new user, the interest label cluster of the old user, and the proportion of the labels used by the old user in each interest label cluster, using formula (6) calculate the interest attribute similarity $sim\_i(u_i, u_j)$ between users;
5. According to the natural attribute similarity $sim\_n(u_i, u_j)$ calculated in step 3 and the interest attribute similarity $sim(u_i, u_j)$ calculated in step 4, use formula (7) to calculate the similarity between users which based on user portrait $sim\_p(u_i, u_j)$;
6. According to the user rating matrix, calculate the similarity $sim\_r(u_i, u_j)$ of users based on service scoring by using formula (12);
7. Calculate the user's direct trust $tru\_d(u_i, u_j)$ and indirect trust $tru\_id(u_i, u_j)$ based on the user's friend relationship, interaction information, and associated path. Use formula (17) to get the user's social trust degree $Tru(u_i, u_j)$;
8. The user portrait similarity and the similarity of scoring obtained according to step 5 and step 6 and using formula (18) to calculate the similarity $sim(u_i, u_j)$ between users;
9. If the target user is an old user, according to the $sim(u_i, u_j)$ and $Tru(u_i, u_j)$ calculated in steps 8 and 7, use formula (19) to obtain the influence $affect(u_i, u_j)$ between users. Let K users with highest influence be the neighboring user set of the target user, and record as G;
10. According to the neighbor user set G, using the formula (20) to obtain the prediction scoring value, make the K services with highest prediction rating value as recommendation service sets recommend to the user, then the recommendation ends;
11. If the target user is a new user, calculate the old user who has the highest similarity to the new user based on the user portrait and make this old user as the neighbor user of the new user;
12. The service recommendation set of the old user among the neighbor users is directly recommended to the new user as a recommendation result, and the recommendation ends.

4 Experimental Results and Analysis

4.1 Data Sources and Preprocessing

This paper selects the Hetrec2011-Last.fm data subset from the Hetrec2011 data set as the experimental data. The data set includes user information, artist information, label information, the number of times the user has listened to the artist, friend-to-friend relationship, the information of timestamp that tag was labeled, etc. These information in the Hetrec2011-Last.fmdata set can be used to well verify the proposed method in this paper.

Firstly, the sample data needs to be preprocessed to remove noise data from the source data, and at the same time it also needs to generate 30% new users to join the data set. In this data set, we take the number of times the user has listened to the artist as the user's rating on the artist. Since the rating in the general data set is a 5-point system, this paper also quantifies the rating to 1 to 5. In this experiment, we divide the data set into training and test sets, in which 80% of the data is placed in the training set and 20% of the data is placed in the test set.

4.2 Evaluation Index

After the recommendation process is over, whether the recommendation result is reasonable needs to be measured by some evaluation indexes. The evaluation index can calculate the accuracy of the predicted value, and can also measure whether it can helps the target users find new points of interest, etc. This paper uses the Mean Absolute Error (MAE) and the Root-Mean-Square Error (RMSE) to evaluate whether the service recommendation result is accurate. These two evaluation indexes can calculate the error between the user's predicted rating for the item and the actual rating, thereby can be used to evaluate the accuracy of the recommendation algorithm. The calculation formulas for MAE and RMSE are given below:

$$MAE = \frac{\sum_{i=1}^{n} | r_{i,k} - p_{i,k} |}{n}.. \tag{21}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (r_{i,k} - p_{i,k})^2}{n}}. \tag{22}$$

In formula (21) and formula (22), $r_{i,k}$ represents the actual rating of the target user u_i on the artist s_k, $p_{i,k}$ represents the predicted rating of u_i on s_k, and n represents the number of sample. The smaller the values of MAE and RMSE shows the service recommendation accuracy is higher.

4.3 Experimental Results and Analysis

4.3.1 Improving the Recommendation Accuracy

The proposed method in this paper takes into account both two situations: the recommendation for old users, and the recommendation for new users. Experiment one is for the first situation, and Experiment two is for the second situation.

Experiment one: This experiment recommends for target users without considering new users and compares with the improved collaborative filtering-based service recommendation method (RACF) [14] and the trustworthy service recommendation method (STSRM) [15]. In this paper, on the basis of the user rating similarity, the user portrait similarity and the social trust degree are added to improve the accuracy of recommendations. The experimental results are shown in Fig. 4 and Fig. 5.

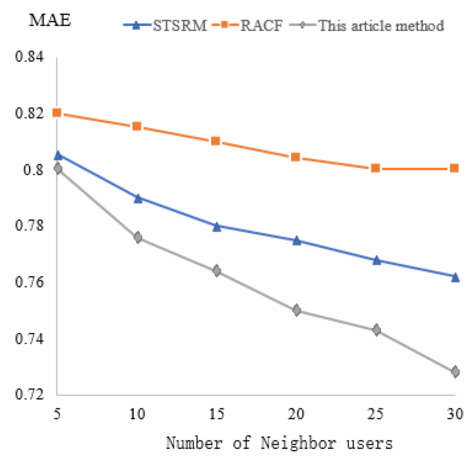

Fig. 4. MAE values of several methods undermalicious recommendation

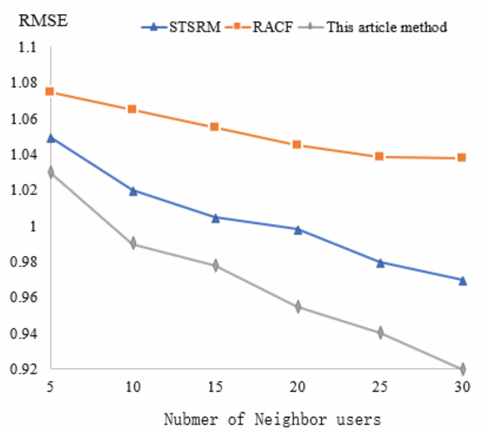

Fig. 5. RMSE values of several methods under malicious recommendation

It can be seen from Fig. 4 and Fig. 5 that as the number of neighbor users continues to increase, the MAE and RMSE values of several recommendation methods have a certain extent decline. This indicates that the more the number of neighbor users, the more accurate the recommendation results. Among them, the RACF method only considers the similarity between users, so the downward trend of MAE and RMSE values is not obvious. This indicates that when there are the malicious recommendation, as the number of neighbor users increases, the improvement of RACF's recommendation result accuracy is not obvious. The STSRM method incorporates the social trust relationship between users on the basis of user similarity, so the downward trend of MAE and RMSE values of STSRM is more obvious than the one of RACF. This indicates that even then there are the malicious recommendation, as the number of neighbor users increases, the improvement of STSRM's recommendation result accuracy is very obvious, and the STSRM's accuracy of recommendation result are also better than the one of RACF.

This proposed method in this paper introduces the concept of timeliness to the calculation of user similarity for improving the accuracy of similarity calculation, and incorporates the social trust degree for resisting the malicious recommendation, then builds the user portraits model, at last combines the user rating similarity and the social trust degree with the user portrait similarity to recommends for target users. Therefore, it can be seen from Fig. 4 and Fig. 5 that the downward trend of MAE and RMSE values of our proposed method is very obvious, and the proposed method has lower MAE and RMSE values than the previous two methods. This indicates that when there are the malicious recommendation, as the number of neighbor users increases, the improvement of recommendation result accuracy of our proposed method is very obvious, and the accuracy of the recommendation of our proposed method also is highest in the three methods.

4.3.2 Alleviating the Cold Start Problem

Experiment two: This experiment is executed by comparing with the improved collaborative filtering-based service recommendation method (RACF) [14] and the trustworthy service recommendation method (STSRM) [15]. In the experiment two, the target users are new users. In order to verify the alleviation ability of our proposed method for the cold start problem, the proportion of new users was continuously increased in the experiment. The alleviation ability of these three methods can be reflected through the changes on their MAE and RMSE values. The experimental results are shown in Fig. 6 and Fig. 7.

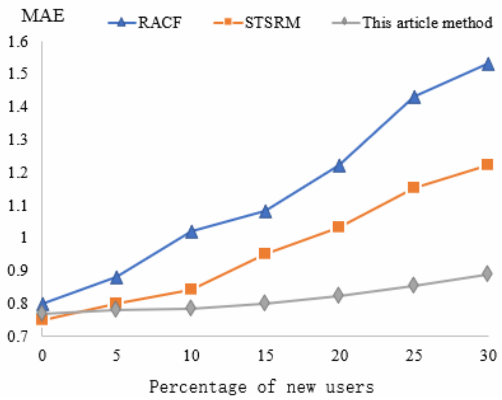

Fig. 6. MAE values of severalmethods

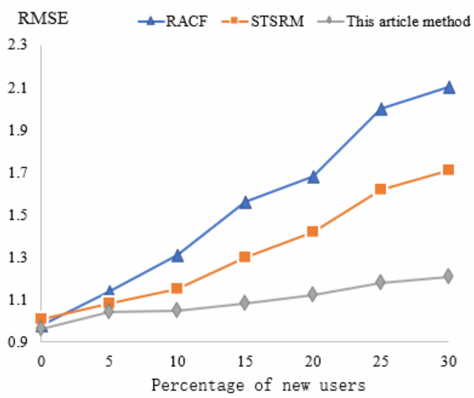

Fig. 7. RMSE values of severalmethods

As can be seen from Fig. 6 and Fig. 7, as the proportion of new users increase, the MAE and RMSE values of several recommended methods have increased to some extent. This indicates that the bigger the proportion of new users, the more inaccurate the recommendation results. Among them, the increase of RACF's MAE and RMSE values is very obvious. This is because of that the RACF method only considers the similarity between users. So it can't recommend well for new users. The increase of STSRM's MAE and RMSE values is not very obvious. This is because of that the STSRM method incorporates the social trust relationship. So it can alleviate the cold start problem for new users to a certain extent.

Compared with the above two methods, the increase of the MAE and RMSE values of our proposed method is not obvious. This is because of that the proposed method introduce the concept of user portrait to the calculation of user similarity, and combines user portrait similarity with the social trust degree. So it can alleviate the cold start problem for new users better, and it also can recommend for new users with the better recommendation result accuracy than the other two method.

5 Conclusion

For the difficulty of users choosing services due to information overload, this paper proposes a Web services hybrid recommendation method based on user portrait. This method mainly considers the two situations of new users and old users, and solves the problem of user cold start well. Firstly, we establish a user portrait model and respectively make portraits for new and old users. The user portrait mainly includes two aspects: natural attributes and interest attributes. Secondly, when calculating the similarity, the user portrait similarity and the user rating similarity are simultaneously considered to improve the accuracy of the similarity calculation. At the same time, the user's social relationship network is considered and a trust calculation model is established that can resist maliciously recommender users well. Finally, we respectively give different recommended results for new and old users. Through experimental comparison, it is found that the proposed method in this paper can improve the accuracy of recommendations and at the same time it can make good recommendations for new users, alleviating the new user cold start problem.

However because the proposed method in this paper has many calculation steps, including the user portrait similarity, the user rating similarity and the social trust degree, the recommendation process is a little bit time consuming, in addition, in order to make the user portrait similarity keep accurate, the constructed user portrait model by clustering needs to keep updating over time. In future work, we will think about how to simplify the complexity of the recommendation process, and at the same time take into account the more user and service feature information to further improve the accuracy of the recommendation method, for example, the user behavior information, the service content information, the knowledge graph information, and so on.

Acknowledgements

Foundation items: Project supported by the National Natural Science Foundation, China (No. 61601146), the Postdoctoral Science Foundation of Heilongjiang Province, China [grant number LBH-Z19067], and the Natural Science Foundation of Heilongjiang Province, China (No. QC2018081).

References

[1] L.-W. Liu, F. Lecue, N. Mehandjiev, Semantic content-based recommendation of software services using context, ACM Transactions on the Web 7(3)(2013) 17-20.

[2] D. Valcarce, J. Parapar, A. Barreiro, Finding and analysing good neighbourhoods to improve collaborative filtering, Knowledge-Based Systems, 2018.

[3] M. He, W.-S. Liu, J. Zhang, Association rules recommendation algorithm supporting recommendation nonempty, Journal on Communications 38(10)(2017) 18-25.

[4] L. Yao, Q. Sheng, H. NIu, Y. Yu, A. Yu, Unified Collaborative and Content-based Web Service Recommendation, IEEE Transactions on Services Computing 8(3)(2015) 453-466.

[5] N. Polatidis, C.-K. Georgiadis, A Multi-Level Collaborative Filtering Method that Improves Recommendations, Expert Systems with Applications 48(2016) 100-110.

[6] H.-G. Rong, S.-X. Hou, C.-H. Hu, J.-X. Mo, User similarity-based collaborative filtering recommendation algorithm, Journal on Communications 35(2)(2014) 16-24.

[7] M. Wang, J. Ma, A Novel Recommendation Approach Based on Users' Weighted Trust Relations and the Rating Similarities, Soft Computing 20(10)(2015) 3981-3990.

[8] P. Bedi, R. Sharma, Trust Based Recommender System Using Ant Colony for Trust Computation, Expert Systems with Applications 39(1)(2012) 1183-1190.

[9] R. Saia, L. Boratto, S. Carta, A semantic approach to remove incoherent items from a user profile and improve the accuracy of a recommender system, Journal of Intelligent Information Systems 47(1)(2016) 111-134.

[10] D. Wang, Research on Personalized Recommendation Algorithm Based on user portrait, Chongqing University, 2018.

[11] Y. Liu, X.-Y. Wu, B.-J, Xie, Research on Information Recommendation Based on Dynamic User Portrait, Computer Systems Applications 27(6)(2018) 236-239.

[12] J.-W. Zhang, N.-B. Wang, S.-B. Huang, S.-M. Man, Research on Bisecting K-Means Clustering Algorithm Optimization and Parallelism, Computer Engineering 37(17)(2011) 23-25.

[13] C. Alan, The Inmates are Running the Asylum, Beijing: Electronics industry publishing house, 2006.

[14] X. Wu, B. Cheng, J. Chen, Collaborative Filtering Service Recommendation Based on a Novel Similarity Computation Method, IEEE Trans. Services Computing 10(3)(2017) 352-365.

[15] W.-Q. Zhu, Y.-S. Zhong, J. Xu, Trustworthy Service Recommendation Method Based on Social Trust, Journal of Chinese Computer Systems 38(3)(2017) 503-508.