Space-Based Information Service Recommendation Algorithm based on Hybrid Strategy

Yeyin Wang¹,a, Shenwei Wang¹, Jie Yu¹,b*, Yufeng Gao¹, Rui Wang¹, Feng Wang¹, and Ying Li¹

¹Institute of information engineering China Academy of Aerospace Systems Science and Engineering, Beijing, 100037, China

a e-mail: wangyeyin2010@163.com, b* e-mail: yy921006@mail.ustc.edu.cn

Abstract. The recommendation technology has been applied to all walks of life, but it is rarely used in space-based information service. Considering the advantages of the recommendation technology and the particularity of space-based information service, a space-based information service recommendation algorithm based on hybrid strategy is proposed. Based on the different stages of the user, the similarity calculation is carried out from three aspects of user attributes, non-common service and service interest. The missing data in the service interest model is predicted by singular value decomposition to solve the problems of cold start and data sparse. Finally, the proposed algorithm is compared with the traditional algorithm, and the accuracy and scalability of the proposed algorithm are improved.

1. Introduction
Space-based information service is the process of delivering space-based information needed by users to users. The existing space-based information service mode is relatively single, and mainly adopts the passive service mode. When more information is available, the passive service lacks flexibility, provides the same information to all users, and does not consider the preferences of users. In the face of massive space-based information, it is more difficult for users to find target information, and the efficiency of obtaining services is low.

Recommendation algorithm is a kind of technology that can help users quickly obtain valuable information. At present, recommender technology has been deeply studied, but space-based information service is different from Internet information service. Space based information has higher requirements in accuracy, security, timeliness and so on. Recommender technology is directly applied in space-based information service.

Considering the advantages of personalized recommendation and the particularity of space-based information service, this paper proposes a space-based information service recommendation method based on hybrid strategy (HS-SISR). Under strict user rights management, the implicit user behavior information is filtered from three aspects of user attributes, non-common service and service interest, and the missing data in the service interest model was predicted by singular value decomposition (SVD) to solve the problems of cold start and data sparse. Finally, the recommendation service for space-based information users is formed.

2. Related work
With the increasing number of users and products, it brings a variety of problems to the traditional recommendation algorithm, such as data sparsity, cold start and low real-time. Therefore, some
scholars have optimized the recommendation algorithm. Wang Zhiyuan\cite{1} et al. improved Slope One algorithm of filling scoring matrix and corrected the results according to users' interests and preferences, effectively solving the problem of traditional collaborative filtering method when data sparsity increases. Suryakant\cite{2} combined cosine similarity, Jaccard coefficient and other three similarity calculation methods to improve the prediction accuracy. Wilson\cite{3} et al. fully considered the global impact of user ratings and utilized all ratings between two users to improve the accuracy of prediction. Liu Qingqing\cite{4} et al. adopted singular value decomposition technology to optimize user similarity and improve the effectiveness of the algorithm.

However, the application of recommendation technology in space-based information service still has some problems as follows:

- **User cold start.** When a new user is registered, there is no history of the user in the system, and the user service interest cannot be calculated.
- **Data sparsity.** When the number of users applying for services in the system is very rare, the prediction accuracy of users' preferences based on joint application service will be reduced.
- **Timeliness.** Too much information provided by users will affect the quality of information service. Implicit information feedback should be adopted as far as possible to avoid disturbing users.
- **Security.** Space-based information services have high requirements for security, and if the user's behavior is described specifically, there will be a risk of information leakage.

3. **The new algorithm model**

This paper presents a hybrid strategy based recommendation algorithm for space-based information services. As shown in figure 1, first, users register or apply for services, and relevant user information and behavior information are recorded without disturbing users. Then, filtering is carried out based on user attributes and behavior information, including service interest model construction, similarity measurement, user preference prediction and sorting of screening results. Finally, according to the characteristics of the user's stage, push the corresponding service content for the user.

![Figure 1. The algorithm flow chart.](image)
3.1. Related Data

3.1.1. User attribute data. User attribute data includes user authority, industry and geographic location. Among them, the user's rights and industry fields are selected when new users register, and the user's geographic location is obtained by the system's embedded positioning. User rights include public users, government users, military users and so on. Industry sectors include agriculture, transportation, environmental protection, military, etc. User geographic location is an important factor to reflect user needs.

3.1.2. User behavior data. User behavior data includes service application time and service application times, which can directly reflect the user's demand for the service.

3.1.3. Service feature data. Service feature data includes geographic location of service target, physical attributes of service target and service load type. When users’ apply for services for many times in a row, there is often a geographical correlation between various services. Physical attribute of service target include mountains, rivers, lakes, buildings, etc. Service load type refers to the performance parameters of sensors used in service.

3.2. Construction of service interest model

Considering that the process of user receiving service is not affected as much as possible, only the number of services in user behavior data is collected, and the number of services is taken as a measure of user interest.

We assume that and $U = \{u_1, u_2, u_3, \ldots, u_M\}$ and $S = \{s_1, s_2, s_3, \ldots, s_N\}$ are the set of users and services, respectively. The data set $A_k$ records the number of times all users call all services, as follows:

$$A_k = \begin{bmatrix} a_{11}(k) & \cdots & a_{1N}(k) \\ \vdots & \ddots & \vdots \\ a_{M1}(k) & \cdots & a_{MN}(k) \end{bmatrix}$$

(1)

where $A_k$ represents the data of the $k$-th period, $a_{us}(k)$ is the number of times user $u$ applies for the service $s$ in the $k$-th period.

The more users apply for the same service in a short time, the higher their interest in the service. User $u$'s interest in service $s$ in the $k$-th period can be expressed by equation (2).

$$a'_{us}(k) = \frac{a_{us}(k) - \min_{1 \leq s \leq N} \left[ a_{us}(k) \right]}{\max_{1 \leq s \leq N} \left[ a_{us}(k) \right] - \min_{1 \leq s \leq N} \left[ a_{us}(k) \right]}$$

(2)

In the $k$-th period of time, the service interest matrix $A'_k$ can be expressed by equation (3).

$$A'_k = \begin{bmatrix} a'_{11}(k) & \cdots & a'_{1N}(k) \\ \vdots & \ddots & \vdots \\ a'_{M1}(k) & \cdots & a'_{MN}(k) \end{bmatrix}$$

(3)

The service interest model for user set $U$ and service set $S$ can be expressed as follows:

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1N} \\ \vdots & \ddots & \vdots \\ r_{M1} & \cdots & r_{MN} \end{bmatrix}$$

(4)

where $r_{us}$ is the degree of that user $u$ is interested in service $s$.

3.3. Similarity calculation based on user attributes

When the user is a new user or the number of service applications is too small, the user attributes are used to find similar users for recommendation, so as to effectively deal with the cold start problem.
The characteristic attributes of user $u$ are represented by $Cus_u = \{c_{u1}, c_{u2}, c_{u3}, \ldots, c_{un}\}$, $n$ represents the number of user characteristic attributes. The attribute similarity of users $u$ and $v$ can be calculated as follows:

$$sim_{Cus}(u,v) = \sum_{c \in Cus} w_c \cdot sim(u,v,c)$$

where $sim_{Cus}(u,v)$ is the similarity of user $u$ and $v$ in attribute $c$, $w_c$ is the weight of user attributes.

3.4. Similarity calculation based on non-common services

In the initial stage of service application, users and other users apply for the same service fewer times. Therefore, the relationship between non-common services is used as a weight factor to adjust the final similarity.

In order to fully evaluate the similarity between users $u$ and $v$, considering the application records of $u$ and $v$ on all services, the cumulative impact of all possible applications on user similarity is calculated as follows:

$$sim_{non-co-rated}(u,v) = \sum_{i \in I_u} \sum_{j \in I_v} sim_{item}(i,j) \cdot sim_{based-item}(r_u, r_v)$$

where $I_u$ represents the service set applied by user $u$, $I_v$ represents the service set applied by user $v$, $sim_{item}(i,j)$ represents the similarity between services $i$ and $j$, and $sim_{based-item}(r_u, r_v)$ represents the similarity between user $u$ and user $v$ in the number of times of applying for service $i$ and $j$, $r_u$ represents the degree of that user $u$ is interested in service $i$, $r_v$ represents the degree of that user $v$ is interested in service $j$.

The corresponding relationship between service and feature attributes is represented by $S_t$ as follows:

$$S_t = \begin{bmatrix} s_{11} & \cdots & s_{1N} \\ \vdots & \ddots & \vdots \\ s_{M1} & \cdots & s_{MN} \end{bmatrix}$$

where $s_{np}$ is whether the $n$-th service has the $p$-th characteristic attribute, 0 or 1.

Similarity between services $i$ and $j$ can be calculated as follows:

$$sim_{new}(i,j) = \frac{\sum_{k=1}^{p} s_{ik} \cdot s_{jk}}{\left(\sum_{k=1}^{p} s_{ik}^2 \sum_{k=1}^{p} s_{jk}^2\right)^{1/2}}$$

3.5. Similarity calculation of service interest

With the increasing number of services applied by users, the similarity calculation is based on service interest.

The premise of Pearson coefficient in calculating user similarity is that two users have a set of common application services, but even if the intersection of two users is very large, there are still gaps in $R_t$, the calculated similarity will be inaccurate. In this paper, SVD\[5\] algorithm is introduced to predict the missing data in the interest matrix of the service interest model, so as to complete the interest prediction, alleviate the problem of data sparsity, and effectively hide the dominant features of user behavior data, which virtually increases the security.

The algorithm process of this stage is as follows:

- The similarity between services is calculated according to equation (8). $R'_t$ is obtained after filling the vacancy of $R_t$. 
• Singular value decomposition of $R'$;
• Select the largest $k$ singular values and solve the user’s nearest neighbor set $U'$;
• Calculate the predicted value of the user's interest in the unapplied service and replace the populated data in the first step to get the complete $R''$;
• Use equation (9) to calculate the similarity between users and $U'$ internal users.

$$sim_{cor-rated}(u,v) = \frac{\sum_{k \in K_{uv}} (r_{uk} - \overline{r})(r_{vk} - \overline{r})}{\left(\sum_{k \in K_{uv}} (r_{uk} - \overline{r})^2 \sum_{k \in K_{uv}} (r_{vk} - \overline{r})^2\right)^{1/2}} \tag{9}$$

As in equation (9), $K_{uv}$ is the set of services that both user $u$ and user $v$ have applied for, $K_{uv} = K_u \cap K_v$.

3.6. Mixed similarity calculation
In this paper, similarity calculation is divided into three stages: user attribute similarity, non-common service similarity and service interest similarity. Mixed-similarity model is used to calculate the final similarity of users, as follows:

$$sim(u,v) = \alpha sim_{	ext{attr}}(u,v) + \beta sim_{	ext{non-cor-rated}}(u,v) + \gamma sim_{	ext{cor-rated}}(u,v) \tag{10}$$

where $\alpha$, $\beta$ and $\gamma$ represent the weights of the three kinds of similarity respectively.

Sigmoid function is introduced to represent the weight $\alpha$, $\beta$ and $\gamma$, which makes the weight conform to the monotonic decreasing characteristic and keep the value between 0 and 1. As the number of users applying for services in the system gradually increases, the weight of similarity in the first stage decreases continuously, while the weight of similarity in the second stage decreases continuously with the increase of the number of jointly applied services. The three weighting factors are defined as follows:

$$\alpha = \frac{1}{1 + \exp\left(-\left|N_u\right|\right)} \tag{11}$$

$$\beta = \frac{1}{1 + \exp\left(-\frac{N_u \cap N_v}{N_u}\right)} \tag{12}$$

$$\gamma = 1 - \alpha - \beta \tag{13}$$

where $N_u$ represents the total number of times that user $u$ applied for service, and $N_v$ represents the total number of times that user $v$ applied for service.

3.7. User Preference Prediction Based on Mixed Similarity
Through the calculation of the above steps, the system can find the user’s nearest neighbor set and predict the user’s interest in a certain service by combining the service application times of the nearest neighbor. The calculation formula is shown in equation (14). Sort the predicted results to generate a recommendation list.

$$r_{us} = \overline{r_u} + \sum_{v \in U'} \frac{sim(u,v)(r_{uv} - \overline{r})}{\sum_{v \in U'} sim(u,v)} \tag{14}$$
4. Experiments

4.1. The data set
The experiment adopts the data set of the satellite remote sensing cloud service platform, including 31882 service records of 428 types of services (geographical locations) recorded by 193 users in 380 days, and 17413 user-service access relationships, with a sparseness of 78.92%. As can be seen, the scoring matrix is very sparse. In this experiment, the data set is divided into 80% training set and 20% test set to analyze the accuracy of service prediction and the scalability of the algorithm. The experimental environment is MATLAB R2016a on Windows 10 professional (64bit), AMD A10-7890K Radeon R7.

4.2. Evaluation standard
In this experiment, classical evaluation indexes mean absolute error (MAE)[6] and algorithm time consumption[7] are used to evaluate the quality of the algorithm. MAE is calculated with equation (15) to measure the accuracy of the prediction. The scalability of the algorithm is evaluated by the time it takes.

\[
MAE = \frac{\sum_{u,i} |r_{ui} - \hat{r}_{ui}|}{n}
\]

As in equation (15), \(r_{ui}\) represents the true situation of user \(u\) for service \(i\), \(\hat{r}_{ui}\) represents the prediction of user \(u\) for service \(i\), and \(n\) represents the number of predictions. The smaller the value of MAE is, the smaller the deviation between the predicted value and the actual value is, and the higher the prediction accuracy is.

4.3. Results and analysis
MAE and algorithm time consumption were selected as metrics, the traditional algorithm UCF and the newly proposed algorithm SVD-HCF were compared with HS-SISR.

Figure 2 shows the variation of MAE when the three algorithms select different number of nearest neighbors. It can be seen that as the number of nearest neighbors increases, the MAE value of each algorithm decreases. When it exceeds a certain value, UCF and SVD-HCF tend to be stable after a certain fluctuation. The MAE value of HS-SISR has a gentle downward trend, which indicates that the algorithm considers the attributes of users' application service at each stage to make predictions, and effectively maintains the prediction accuracy.

Figure 3 shows the time consuming statistics of the three kinds of algorithms in the case of selecting different numbers of nearest neighbors. It can be seen that the time consumption of each algorithm increases gradually. When the number of neighbors exceeds 60, the time consumption of SVD-HCF and HS-SISR increases slowly and is lower than that of UCF. The time consumption of HS-SISR is similar to that of SVD-HCF. However, the time consumption of HS-SISR changes slowly after N exceeds 20. When N exceeds 80, the time consumption is significantly lower than that of SVD-HCF, indicating that the HS-SISR proposed in this paper has better performance and strong scalability.
5. Conclusions

Facing the special requirements of timeliness and security for space-based information service recommendation, this paper proposes a space-based information service recommendation algorithm based on hybrid strategy to alleviate the problems of user cold start and data sparsity. The experiments show that the user has good prediction accuracy in the initial stage and the data sparsity stage, and the time consumption of the algorithm increases slowly and has certain scalability. Although the research in this paper has achieved certain stage results, it does not consider that users' interest will gradually fade and be forgotten as time goes by. The next step is to consider the influence of time factor on the dynamic change of users' interest, so as to improve the accuracy of the recommendation system.

Acknowledgments

This paper is one of the phased achievements of IAA's research topic "Global Satellite Data Sharing Mechanism".

References

[1] Wang, Z.Y., Wang. X.F. (2020) A personalized recommendation algorithm based on improved matrix filling of user interest difference. Computer Applications and software, 37(12): 224-230+237.

[2] Zhang, P.F., Zhang, Z.J., Tian, T., et al. (2019) Collaborative filtering recommendation algorithm integrating time windows and rating predictions. Applied Intelligence, 49(8): 3146-3157.

[3] Wilson, J., Chaudhury, S., Lall, B. (2014) Improving Collaborative Filtering Based Recommenders Using Topic Modelling. In: 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). ACM.

[4] Liu, Q.Q., Luo, Y.L., Wang, Y.F., Zheng, X.Y., Chen, W. (2019) Hybrid recommendation algorithm based on SVD filling. Computer Science, 46(S1): 468-472.

[5] Cui, L.Z., Huang, W.Y., Qiao, Y., et al. (2017) A novel context-aware recommendation algorithm with two-level SVD in social networks. Future Generation Computer Systems, 86(SEP.):1459-1470.

[6] Wei, Z., Zhou, D. (2017) Optimization collaborative filtering recommendation algorithm based on ratings consistent. In: IEEE International Conference on Software Engineering & Service Science. IEEE.

[7] Pang, Y.J., Jia, S.J. (2018) Design of Intelligent Monitoring System for Hazardous Chemicals Based on Internet of Things Technology, Chemical Engineering Transactions (CET Journal), 71(5).