2021 3rd International Conference on Advances in Computer Technology, Information Science and Communication (CTISC) | 978-1-6654-1868-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/CTISC52352.2021.00055

A Collaborative Recommendation Model of Agricultural Planting Technology Based on User Characteristics

Yinfan Hui
College of Information Engineering
Northwest A & F University
Yangling, Shaanxi, China
yinfan@nwafu.edu.cn

Shuqin Li*
College of Information Engineering
Northwest A & F University
Yangling, Shaanxi, China
lsq_cie@nwsuaf.edu.cn

***Abstract*—In view of the lack of informatization in the field of agricultural planting production, the traditional collaborative filtering recommendation algorithm has problems such as cold start, sparse scoring matrix, and poor scalability, resulting in poor recommendation quality. A personalized recommendation model of collaborative filtering agricultural planting technology that integrates user characteristics is proposed. First, the user's initial characteristics are constructed according to the user's geographic location, planting occupation, and main crops, and user behavior information is used to update user characteristics. Then combine the user characteristic similarity model and the rating matrix similarity model, and reconcile the weighting factors to form the user's comprehensive similarity. Finally, use Top-N for personalized recommendation. By integrating user characteristics, the recommendation model is more suitable for agricultural planting scenarios, and the recommendation results are more flexible and reasonable. Experimental results show that compared with user-based collaborative filtering recommendation and user-characteristic-based recommendation, the proposed algorithm improves precision by 2% and 7%, recall rate by 2% and 9%, and F1 value is 71%. The effectiveness of the proposed method is verified.**

Keywords—User characteristics, collaborative filtering, planting technology, personalized recommendation, similarity

I. INTRODUCTION

The output value of China's planting industry accounts for more than 50% of the total agricultural output value. It is an important part of agriculture. According to the current problems in the field of agricultural planting production, such as low technological innovation ability and low agricultural production efficiency[1]. Therefore, planting technology recommendation has become an effective way for farmers to obtain information and improve agricultural production efficiency. Personalized agricultural planting technology recommendation has become a practical problem that needs to be studied.

Collaborative filtering recommendation algorithm[2] is the most widely used recommendation algorithm in personalized recommendation technology. However, it considers a single factor, and has the problems of user cold start, sparse data and poor scalability[3]. In order to solve the above problems, researchers have proposed a variety of solutions, such as hybrid recommendation strategies, trust mechanisms and so on. In [4], introduce user basic information (including gender, age, location) and social information (user preferences and social friends), and integrate user trust and user attribute similarity in the collaborative filtering algorithm. In [5], the ungraded items in the matrix are filled with the predicted value of the score calculated by the slope-one algorithm to alleviate the data sparseness. In [6], the collaborative filtering part uses dimensionality reduction technology singular value decomposition. Find the most similar users and items in each user and item cluster, so as to improve the scalability of the recommendation algorithm.

Inspired by the above method, combined with the characteristics of agricultural planting technology information recommendation. A collaborative filtering and personalized agricultural planting technology recommendation model based on user characteristics (UC-CF) is proposed. The similarity between users is calculated from multiple angles and methods. The influence of the natural regionality of the agricultural planting scene and the natural regionality of the farmers on the recommendation quality is comprehensively considered. At the same time, the recommended effect is more suitable for agricultural planting.

II. PERSONALIZED RECOMMENDATION MODEL OF AGRICULTURAL PLANTING TECHNOLOGY

According to the diversity, regional and dynamic characteristics of agricultural planting technology information recommendations. This section describes the construction of user characteristics to the recommendation of collaborative filtering agricultural planting technology that integrates user characteristics.

A. Construction of agricultural user characteristics

User characteristic construction is the process of tagging users. The weight of the label represents the degree of user preference.

In the absence of new user behavior data, the user's initial characteristics are firstly constructed in multiple dimensions with user attribute information. The user characteristic labeling system constructed is mainly from the following three dimensions: geographic location, planting occupation, and main planting crops. Then update the user characteristics by collecting the different

978-1-6654-1868-3/21/$31.00 ©2021 IEEE
DOI 10.1109/CTISC52352.2021.00055

behaviors of the user on the planting technology. The behavior operation includes the user's browsing, liking, collecting, forwarding, commenting on the planting technology and other information that reflects the user's planting preference. Therefore, the user characteristic label system of the personalized recommendation system for agricultural planting technology constructed in this paper is shown in Figure 1.

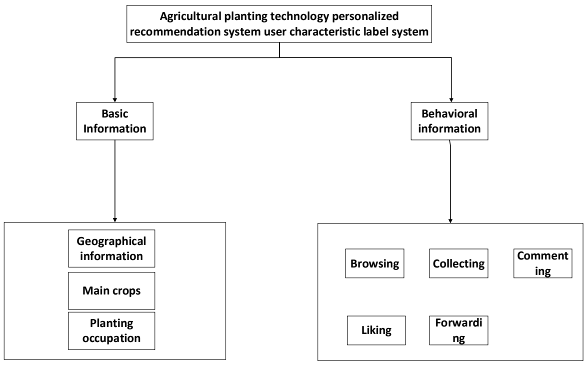

Figure 1. AGRICULTURAL USER CHARCTERISTIC LABEL SYSTEM

User characteristic label weight calculation is the core step of constructing and updating user characteristic labels. This paper divides the calculation of user characteristic tag weight into subjective weight and objective weight. The subjective weight calculation is affected by the behavior type weight and the time of the behavior operation. Different behaviors such as user browsing, liking, collecting, commenting, and forwarding are of different importance to users. Some user behaviors are constantly weakened by the influence of time. The farther the behavior time is from now, the less significant the behavior is for the user. The calculation of the objective weight of behavior is affected by the weight of the label and the number of behaviors. The tag weight is the weight of the tag in the text calculated by TF-IDF. The number of behaviors represents the number of behaviors a user has with the tag on a certain day. The more the number, the greater the impact of the tag on the user. The specific weight calculation method is shown in Figure 2.

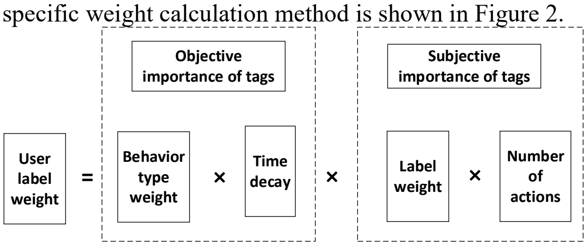

Figure 2. USER LABEL WEIGHTING CACULATION

The calculation of the label weight in the text is based on the collected planting technology title and text content. Use Jieba word segmentation for word segmentation to get the characteristic phrases that represent the main content of the text. Then use the TF-IDF algorithm to extract keywords and calculate the corresponding weights. In order to better segment the agricultural planting technology text, import a custom agricultural dictionary and stop word database before using Jieba segmentation. And cluster the characteristic phrases through K-means. The characteristic words with the highest frequency are selected to represent their characteristic groups and used as user characteristic tags.

According to the calculation formula of the weight of the user characteristic label, the weight of the user characteristic label is obtained. The first n characteristic words with the largest weight are selected as user characteristic tags.

B. Personalized recommendation model integrating user characteristics and collaborative filtering

The collaborative filtering personalized recommendation algorithm (UC-CF) that integrates user characteristics is essentially combined with the characteristics of agricultural planting technology recommendation, and combines two similarity models to alleviate the problems of matrix sparseness, cold start, and poor scalability.

The algorithm includes two aspects. On the one hand, based on the constructed user's personal characteristics tags. First use k-means clustering to form groups with similar preferences to reduce the search time of nearest neighbor users. Then, according to the user characteristic tags, the similarity between users is calculated by the method of Jaccard similarity measurement. On the other hand, filter users in the same region, and filter out users with no scoring items. The ungraded items in the matrix are filled with the predicted value of the score calculated by the slope one algorithm. According to the filled scoring matrix, the similarity between users is calculated using the corrected cosine similarity. Finally, by adjusting the MAE value, determining the best weighting factor and neighbor value, the two user similarities are linearly weighted and fused to generate Top-N recommendations. The process of collaborative filtering personalized recommendation algorithm fusing user characteristics is shown in Figure 3.

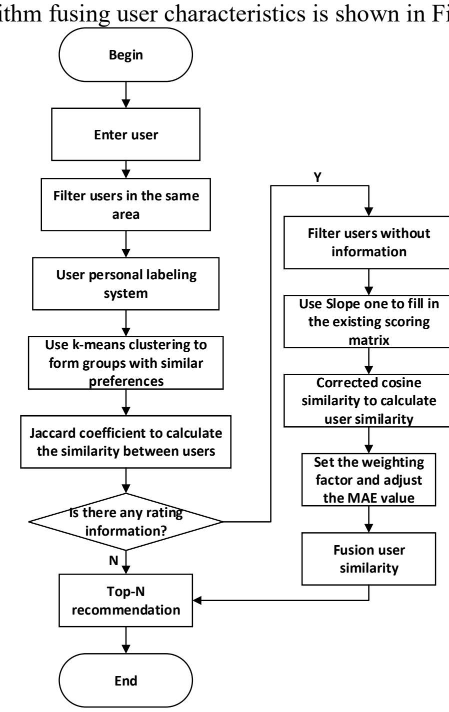

Figure 3. PERSONALIZED RECOMMENDATION ALGORITHM OF AGRICULTURAL PLANTING TECHNOLOGY

Recommendation based on user characteristics (UC) uses the Jaccard similarity measurement method to calculate user similarity, as shown in formula (1).

$$sim_{uc}(u,v) = \frac{|I(u) \cap I(v)|}{|I(u) \cup I(v)|} \quad (1)$$

In the formula, $I(u)$ is the label set of user u. $I(v)$ is the label set of user v.

User-based collaborative filtering recommendation (CF) uses the modified cosine similarity method to calculate user similarity, as shown in formula (2).

$$sim_{cf}(u,v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{vi} - \bar{r}_v)}} \quad (2)$$

In the formula, r_{ui} and r_{vi} represent the ratings of user u and v on item i . $\bar{r}_u$, $\bar{r}_v$ represent the average value of ratings of all items by user u and v; I_{uv} represents the set of items scored by users u and v together.

The two similarity models are weighted to propose a personalized recommendation algorithm (UC-CF) that combines user characteristics and collaborative filtering, as shown in formula (3).

$$sim(u,v) = (1-\lambda) sim_{uc}(u,v) + \lambda sim_{cf}(u,v) \quad (3)$$

In the formula, the interval range of λ is [0,1], which refers to the weighting factor and the similarity fusion parameter. The goal is to balance the similarity of the two models. If the λ value is between 0-1, it is necessary to comprehensively consider the user similarity calculated by different methods and different dimensions to find the nearest neighbor set of the target user to generate Top-N recommendations for the user.

III. EXPERIMENT AND ANALYSIS

A. Experimental data

This study takes farmers in various urban areas of Shaanxi Province as the preliminary research objects. Based on the agricultural data of the Shaanxi Statistical Yearbook issued by the Department of Agriculture and Rural Affairs of Shaanxi Province in 2020. According to each city (district) planting crops and planting area, as well as regional characteristic crops. According to the different species, the agricultural planting professions are preliminarily divided into grain farmers, vegetable farmers, cotton farmers, fruit farmers, melon farmers, bean farmers, tea farmers, and others. According to the classification of planting occupations, the main crops planted by the farmers in each city (district) are determined by means of interview surveys, network searches, and data access. Finally, user-related data is composed of user basic information, user-planting technology rating data, and user behavior information. Therefore, the acquired and expanded data has reference, relevance, and regularity with users.

The data set includes 17,556 scoring data and 68,555 behavioral data on 919 planting information of 120 users. The sorted data set is divided into training and testing data sets at a ratio of 60% and 40%, and a 5-fold cross-validation method is used to make predictions.

Basic user information includes user ID (id), name (username), gender (sex), age (age), main crops (maincrops), occupation (occupation), location(location), etc. The intercepted part of the data is shown in Figure 4.

id	username	sex	age	maincrops	occupation	location
1	yulin01	male	46	corn	grain farmer	Yulin, Shaanxi
7	yulin07	Female	51	soybeans	bean farmer	Yulin, Shaanxi
11	baoji01	male	42	wheat, corn	grain farmers	Baoji, Shaanxi
23	xian03	male	46	peppers, cucumbers, tomatoes, lettuce, leeks	vegetable farn	Xi'an, Shaanxi
25	xian05	male	47	apples, grapes, pomegranates	fruit farmer	Xi'an, Shaanxi
32	yangling02	male	52	peppers, cucumbers, tomatoes, lettuce, leeks	vegetable farn	Yangling, Shaanxi
34	yangling04	male	53	apples, grapes, kiwis	fruit farmer	Yangling, Shaanxi
60	hancheng10	male	46	wheat, corn, rapeseed, pepper	other	Hancheng, Shaanxi
77	yanan07	male	53	soybeans	bean farmer	Yan'an, Shaanxi
81	tongchuan01	male	42	corn, wheat	grain farmers	Tongchuan, Shaanxi
91	hanzhong01	male	42	rice, corn, wheat	grain farmers	Hanzhong, Shaanxi
101	ankang01	male	42	corn, wheat, rice	grain farmers	Ankang, Shaanxi
106	ankang06	male	43	strawberry, citrus, papaya	fruit farmer	Ankang, Shaanxi
120	shangluo10	male	39	corn, wheat, soybeans, rapeseed, peanuts	other	Shangluo, Shaanxi

Figure 4. SOME EXAMPLES OF BASIC USER INFORMATION

User-planting information scoring data includes scoring ID (id), scoring (mark), scoring time (create time), planting information ID (news id), and user ID (user id).

User behavior information scoring data includes behavior ID (id), behavior type (behavior type), operation time (create time), planting information ID (news id), and user ID (user id). Among them, the behavior types are divided into 5 categories, which respectively represent browse, like, favorite, comment, and forward.

B. Experimental environment and metrics

The experimental environment in this paper uses Windows 10 operating system, the processor is Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz, the memory is 8GB, PyCharm2018.3 is used as the development tool, and Python3.6 is used as the development language.

In this paper, the evaluation criteria for the selection of test thresholds and the adjustment of weighting factors adopt the average absolute error (MAE) commonly used in recommendation algorithms. The MAE value is used to indicate the deviation between the user's predicted score and the actual score. The smaller the MAE value, the higher the prediction precision. Define $\hat{R}_{ij}$ to represent the predicted score of user i for item j, R_{ij} is the actual score, and n is the size of the test data set, then the MAE definition is as shown in formula (4).

$$MAE = \frac{\sum_{i=1}^{n} |\hat{R}_{ij} - R_{ij}|}{n} \quad (4)$$

When verifying the comparison of experimental results of different methods, this article uses offline experiments to generate a recommendation list for users. This type of recommendation is called Top-N recommendation. Therefore, this article uses the precision rate (Precision) and recall rate (Recall) to measure. The calculation method of the precision of the recommended results is shown in formula (5).

$$Precision = \frac{\sum_{u \epsilon U} |R(u) \cap T(u)|}{\sum_{u \epsilon U} R(u)} \quad (5)$$

The method of calculating the recall rate of the recommended results is shown in formula (6):

$$Recall = \frac{\sum_{u \epsilon U} |R(u) \cap T(u)|}{\sum_{u \epsilon U} T(u)} \quad (6)$$

In the formula, $R(u)$ is the recommended content list given by the algorithm according to the user training set.$T(u)$ is the content list of the user in the test set.

The Precision and Recall indicators sometimes have contradictions, so need to consider them comprehensively. F1-Measure is the weighted harmonic average of precision and recall. It is the result of combining these two indicators. When the F1 value is high It can show that the test method is more effective. The calculation method is shown in formula (7).

$$F1 - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \qquad (7)$$

C. *Experimental results and analysis*

1) *Weighting factor adjustment*

First, verify the influence of the weighting factor λ on the recommendation result, respectively fix the number of neighbors to k=5, k=10, k=15, and k=20, and compare the changes of MAE under different weighting factors λ. The experimental results are shown in Figure 5. It can be found that MAE first decreases and then increases with the continuous increase of λ, reaching the lowest point between 0.6 and 0.9.

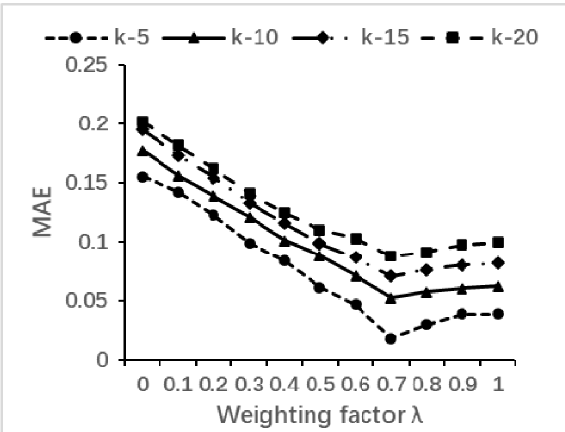

Figure 5. COMPARISION CHART OF MAE VALUES FOR DIFFERENT λ

2) *Neighbor selection*

Continue to verify the influence of the number of neighbors k on the results. The fixed weighting factor λ is 0.6, 0.7, 0.8, 0.9, so that k gradually increases from 2, 3, 5, 10, 15, 20, and 25. The experimental results are shown in Figure 6. When k is in the range of 2 to 3, MAE keeps decreasing. The lowest point is when k=3. As k increases later, MAE continues to increase.

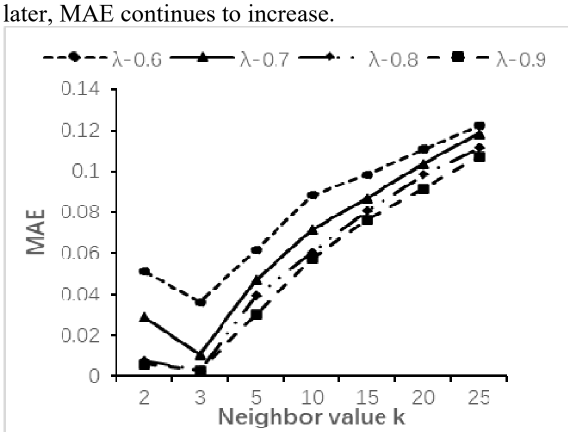

Figure 6. COMPARISION OF MAE VALUES UNDER DIFFERENT NEAREST NEIGHBOUR k

The experimental results show that the MAE value reaches the minimum when λ=0.7 and k =3. So this parameter value is used for subsequent experimental research.

3) *Comparative analysis of experimental results of different methods*

In order to verify the performance of the collaborative filtering personalized recommendation algorithm (UC-CF) proposed in this paper that integrates user characteristics. The precision, recall, and F1 values under different recommended planting techniques were used as evaluation criteria. When the number of recommended planting techniques is 30, 40, 50, 60, 70, and 80, a comparison experiment is carried out with the recommendation algorithm based on user characteristics (UC) and the collaborative filtering recommendation algorithm based on users (CF). As shown in Figures 7, 8, and 9.

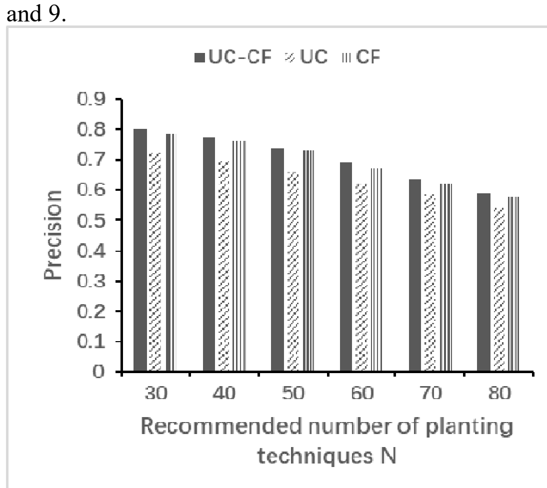

Figure 7. COMPARISION OF THE PRECISION OF THE THREE ALGORITHM

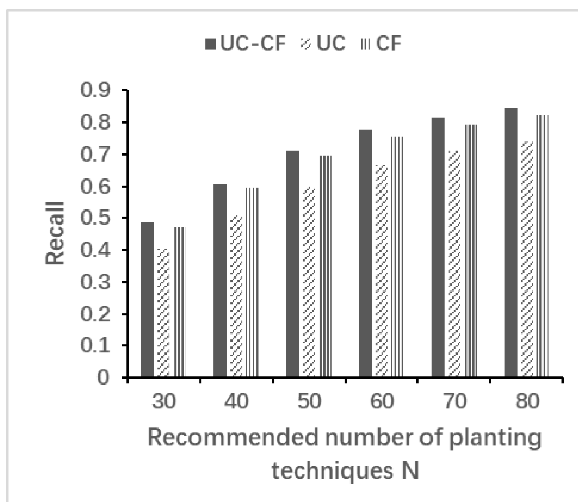

Figure 8. COMPARISION OF THE RECALL OF THE THREE ALGORITHM

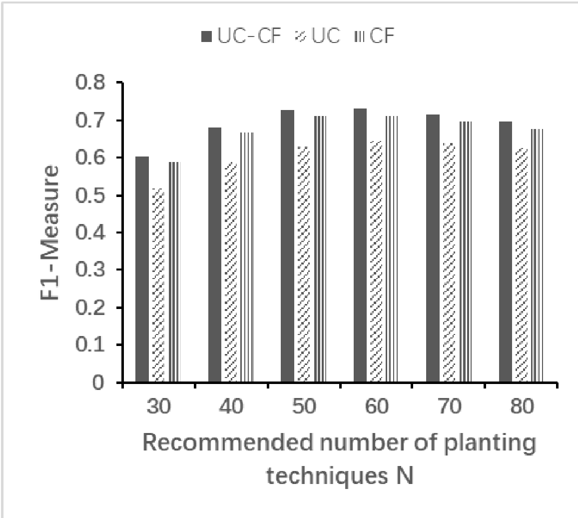

Figure 9. COMPARISION OF THE F1 VALUES OF THE THREE ALGORITHM

From the figure, the precision shows a decreasing trend with the increase of the number of recommendations, and the recall shows an increasing trend with the increase of the number of recommendations. The F1 value increases first and then decreases as the recommended number increases. When the recommended number is 60, the recommendation effect is the best. And UC-CF algorithm is superior to UC and CF algorithm in three dimensions. Therefore, the hybrid algorithm model proposed in this paper can improve the quality of agricultural planting technology recommendation service to a certain extent.

IV. CONCLUSION

In this paper, the recommendation algorithm of single-factor collaborative filtering is not high in the recommendation of agricultural planting technology. There are problems such as sparse scoring matrix, cold start, and poor scalability. A personalized recommendation model of collaborative filtering agricultural planting technology that integrates user characteristics is proposed. This method integrates user characteristic tags. Construct and update user characteristic tags from the attribute information and behavior information of farmers, and calculate user similarity through Jaccard similarity. At the same time, slope One is used to improve the user-based collaborative filtering recommendation algorithm, and the user similarity is calculated through the revised cosine similarity. Finally, the two similarity values are linearly weighted and fused, and the weights are reconciled to obtain the user's comprehensive similarity. It is also used for the selection of nearest neighbors and the recommendation of agricultural planting techniques. This model alleviates the impact of users' cold start, sparse matrix, and poor scalability on the recommendation effect, and improves the quality of recommendation.

Experiments show that the personalized recommendation model of collaborative filtering agricultural planting technology that integrates user characteristics proposed in this paper improves the recommendation effect. However, due to the lack of user-related data, the weighting factor adjustment and nearest neighbor value selection in this article are targeted and dependent on the data. In the following research, more user data sets can be used to select and adjust the parameters in the actual planting scene to make the model more flexible and rigorous.

ACKNOWLEDGMENT

We are grateful for the reviewers' hard work and constructive comments. This work was supported by National Key R&D Program Special Projects (2020YFD1100601).

REFERENCES

[1] Jia L, Wang J, Liu Q, et al. “Application Research of Artificial Intelligence Technology in Intelligent Agriculture,” International Conference on Computer Engineering and Networks. Springer, Singapore, vol.1247, pp.219-225, October 2020.

[2] Linden G, Smith B, York J. “Amazon. com recommendations: Item-to-item collaborative filtering,” IEEE Internet computing, vol.7, pp.76-80, January 2003.

[3] Zhang Yujie, Dong Zheng, Meng Xiangwu. “Research on personalized advertisement recommendation system and its application,” Chinese Journal of Computers, vol.44, pp.531-563, February 2021.

[4] Sedhain S, Sanner S, Braziunas D, et al. “Social collaborative filtering for cold-start recommendations,” Proceedings of the 8th ACM Conference on Recommender systems. pp.345-348, October 2014.

[5] Saeed M, Mansoori E G. “A new slope one based recommendation algorithm using virtual predictive items,” Journal of Intelligent Information Systems, vol.50, pp.527-547, June 2018.

[6] Nilashi M, Ibrahim O, Bagherifard K, “A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques,” Expert Systems with Applications, vol.92, pp. 507-520, February 2018.