

## Research Article

# An Improved Recommendation Method Based on Content Filtering and Collaborative Filtering

Lei Fu <sup>1</sup> and XiaoMing Ma <sup>2</sup>

<sup>1</sup>School of Management Engineering, Tianjin Modern Vocational Technology College, Tianjin 300350, China

<sup>2</sup>School of Intelligent Manufacturing, Tianjin Sino-German University of Applied Sciences, Tianjin 300350, China

Correspondence should be addressed to XiaoMing Ma; maxiaoming@tju.edu.cn

Received 1 March 2021; Revised 10 May 2021; Accepted 21 May 2021; Published 29 May 2021

Academic Editor: Wei Wang

Copyright © 2021 Lei Fu and XiaoMing Ma. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the popularization of the Internet and the prevalence of online marketing, e-commerce systems provide enterprises with unlimited display space and provide customers with more product choices, while its structure is becoming increasingly complex. The emergence and application of the network marketing recommendation system have greatly improved this series of problems. It can effectively retain customers, prevent customer loss, and increase the cross-selling volume of the e-commerce system. However, the current network marketing recommendation system is still immature in practical applications, and the problem of data sparseness is serious. The problem of user interest drift is not well dealt with, resulting in poor recommendation quality and poor real-time recommendation. Therefore, this paper proposes an online marketing recommendation algorithm based on the integration of content and collaborative filtering. First, content-based methods are used to discover users' existing interests. After that, the mixed similarity model of content and behaviour is used to find the similar user group of the target user, predict the user's interest in the feature words, and discover the user's potential interest. Then, the user's existing interest and potential interest are merged to obtain a user interest model that is both personalized and diverse. Finally, the similarity between the marketing content and the fusion model is calculated to form a set of user ratings combined with characteristics and then clustered through K-means to finally achieve recommendation. Experiments have proved that this method has good recommendation performance.

## 1. Introduction

With the rapid development of Internet technology, there are more and more information on the Internet, making it difficult for users to select the information they are interested in. For this reason, a personalized recommendation system came into being to recommend relevant information for users from the Internet [1–3].

At present, personalized recommendation technologies are mainly divided into two types: collaborative filtering [4] and content-based [5]. Collaborative filtering recommendation technology can be divided into user-based and item-based recommendation technology [6, 7]. User-based collaborative filtering recommendation technology predicts item ratings based on the ratings of other users to generate item recommendations [8]. However, its

recommendation quality is easily affected by the sparseness of user evaluation data. The content-based recommendation technology is to analyse the characteristics of the item content information and calculate the matching degree with the user's interest to recommend items [9]. Therefore, compared with collaborative filtering recommendation, content-based recommendation is less dependent on scoring data. However, it has high requirements for the structure and feature extraction of item information, and the recommended items are usually frequently recommended items, which cannot adapt to the recommendation of new items. In view of the respective shortcomings of the two recommendation technologies, some scholars combine the two technologies. Literature [10] used a weight to integrate the scores based on collaborative filtering and content recommendation and make recommendations, to

play their respective advantages. However, it needs to adjust this weight value, and no reasonable adjustment mechanism is given. Literature [11] first used content-based recommendation technology to predict user ratings and builds an initial prediction error matrix. Then, collaborative filtering is used to supplement and perfect the values in the matrix and finally make a final prediction score based on this matrix. Literature [12] first generated multiple preliminary recommendation items through collaborative filtering. Then, the initial recommended item set is deleted through content recommendation technology, and the most relevant recommended items are finally obtained.

To solve the above problems, this paper proposes a fusion recommendation method based on content and collaborative filtering. In this paper, the user's existing interest and potential interest are fused to obtain a user interest model that is both personalized and diverse. By calculating the similarity between the marketing content and the fusion model, a user rating set combining the characteristics is constructed. More accurate recommendation information can be obtained. The main contributions of this paper are as follows:

- (1) This method improves the traditional content-based method to obtain the user's existing interest and obtains the user's potential interest through collaborative filtering of feature words.
- (2) Moreover, the user's existing interest and potential interest are merged to obtain a fused user interest model. The fusion model is used to calculate the similarity of the candidate marketing content and recommend content that may be of interest to different users.
- (3) Compared with the previous method, this article takes into account the diversity and personalization needs of users' browsing products and effectively avoids the time lag of the hybrid recommendation method.

## 2. Related Works

The emergence of e-commerce personalized recommendation is the first realization in the late 1990s with the rapid development of e-commerce. In recent years, the development and innovation of this technology have also been continuously improved. This development and improvement have greatly subverted consumers' traditional consumption patterns, network marketing, and application patterns.

In order to achieve better recommendation effect, many scholars have been focusing on various personalized recommendation design schemes for many years. Among them, the collaborative filtering-based personalized recommendation technology is the most widely used in many personalized recommendations. Kastner et al. [13] developed a system based on collaborative filtering technology, whose main purpose is to filter emails. Goyani et al. [14] developed Group Lens, which is mainly used for collaborative filtering

in newsgroups. Its success has greatly promoted the rapid development of collaborative filtering technology in personalized recommendations. Li et al. [15] applied collaborative filtering technology to build a movie recommendation website Movie Lens. At present, many people use this data set to test and analyse their own algorithms. Tan and He [16] proposed to apply the principle of collaborative filtering to product recommendation. Jiang et al. [6] proposed to introduce the trust model into the collaborative filtering algorithm to increase the accuracy of recommendation. Since then, more and more scholars have proposed various personalized recommendation algorithms based on collaborative filtering, such as personalized recommendation based on adaptive collaborative filtering [17], collaborative filtering based on social psychology [18], collaborative filtering personalized recommendation based on trust awareness [19], and so on.

Currently, collaborative filtering algorithms have been widely used. However, due to the very small proportion of the number of goods purchased to the total number of goods, coupled with the rapid development of the Internet, the number of users and commodities in the recommendation system is very large. The user-project scoring matrix is not only high dimensional but also sparse. This leads to problems such as low timeliness, low precision of recommendation, and cold start of new projects [20]. Aiming at the problem of low solidity, Borlea et al. [21] proposed to solve the problem of "local optimization leading to the sensitivity of initial cluster centres in the k-means partitioning clustering algorithm. Bhattacharjee and Mitra [22] aimed at k-means partitioning. The algorithm can only query the problem of clusters of balls. It is proposed to combine the k-means partition-clustering algorithm with the density-based clustering algorithm. Huang et al. [23] proposed to rely on dynamic search to determine the value of  $k$  autonomously. Nevertheless, this scheme has one drawback that is it is not easy to judge normal clusters and abnormal clusters during the clustering process and is prone to deviation. Ren et al. [24] proposed an improved k-means algorithm, which overcomes the k-means division of clustering. The attribute of the algorithm data is limited. Idrees and Al-Yaseen [25] proposed an algorithm based on the genetic algorithm to find the initial clustering centre. Kolaja et al. [26] proposed multiple subset solutions in the data set for the problem of "local" optimization. Aiming at the problem of data sparsity, Wu and Li [27] introduced dimensionality reduction based on singular value decomposition to optimize data sparsity and decomposed the high-dimensional user-item rating matrix into a low-dimensional orthogonal matrix to better solve the problem of sparsity. Richa and Bedi [28] proposed to use the most frequent score value to predict and fill the sparse score matrix to alleviate the impact of data sparsity.

In response to these problems, some studies have combined the two technologies to propose a hybrid recommendation technology, and related research results have been shown [29, 30]. The hybrid recommendation technology has higher recommendation accuracy than the previous two recommendation technologies.

### 3. Improved Network Marketing Recommendation Algorithm Based on Content Filtering and Collaborative Filtering

**3.1. Recommended System.** A complete recommendation system follows the data input-algorithm processing-data output model, which is mainly composed of input module, recommendation algorithm module, and output module. Each module has its specific function and role and cooperates with each other to complete the recommendation. Its architecture is shown in Figure 1.

- (1) *Data Layer.* The main function of the module is to make full use of different channels to collect and update user information and to provide a channel for the recommendation system and user interaction, which is the basis of the entire recommendation system. Data sources are mainly divided into individual customers and community groups, and feedback information is divided into displayed information and implicit information.
- (2) *Logic Layer.* The module is the core part of the entire recommendation system. The task is to analyse and process the information collected by the data layer module.
- (3) *Business Layer.* The function of the module is to return recommended products to the corresponding users after forming recommended results in different forms. Timeliness and friendliness must be achieved, while ensuring the diversity of output methods.

Figure 2 shows the construction process of the fusion model proposed in this paper. Since EUIM is derived from the direct behaviour of users, its characteristic words are texts traditionally read or written by users, reflecting users' interests and preferences. PUIM uses collaborative filtering to extract feature words that are followed by similar users but the target user has not paid attention to, thus reflecting the potential interest of the target user. FUIM is the result of the integration of the two, which can take into account both the existing and potential interests of users.

**3.2. Existing Interest Model Construction.** Given the product set  $S = \{s_1, s_2, \dots, s_p\}$  and the main feature word sequence  $M = \{m_1, m_2, \dots, m_q\}$ , then the product  $s_i$  can be expressed as a vector space model  $s_i$  corresponding to the main feature word sequence  $a = (a_{i1}, \dots, a_{ij}, \dots, a_{iq})$ . Among them,  $a_{ij}$  represents the weight of the feature word  $m_j$  in the product  $s_i$  and  $a_{ij} = 0$  means that the feature word  $m_j$  has not appeared in the product  $s_i$ , so the entire product set can be expressed as a weight matrix:

$$\text{smt} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1q} \\ \dots & \dots & \dots & \dots \\ a_{p1} & a_{p2} & \dots & a_{pq} \end{pmatrix}. \quad (1)$$

This article uses the TF-IDF notation. Due to the large gap between the lengths of the product text, some text

content is particularly large and some have a few words. In order to prevent long text terms from getting higher weight, formula (2) is used to calculate  $a_{ij}$ :

$$a_{ij} = \frac{\text{count}(i, j)}{\max\_count(i, j)} * \lg\left(\frac{N}{\text{Num}(j)}\right). \quad (2)$$

Among them,  $\text{count}(i, j)$  is the number of times the feature word  $j$  appears in product  $i$ ,  $\max\_count(i, j)$  is the maximum number of times other feature words appear in product  $i$ ,  $N$  is the total number of products, and  $\text{Num}(j)$  is the appearance.  $j$  is the number of products passing feature word.

User interests usually change over time. In this regard, this article proposes a time-weighted EUIM calculation method, which is combined with the time the user clicks to generate EUIM.

Suppose user  $u$  browses the product set  $Su = (su_1, \dots, su_l)$ , where the time of browsing product  $sui$  is  $t_{sui}$  and the current time is  $t$ , then the time influence factor  $\lambda$  of product  $sui$  on user  $u$  is defined as follows:

$$\lambda(sui, u) = \frac{e^{-(t-t_{sui})}}{\sum e^{-(t-t_{sui})}}, \quad (3)$$

where  $Su$  is a subset of  $S$  and its weight matrix is as follows:

$$\text{smt}_{su} = \begin{pmatrix} a_{u11} & a_{u12} & \dots & a_{u1q} \\ \dots & \dots & \dots & \dots \\ a_{up1} & a_{up2} & \dots & a_{upq} \end{pmatrix}, \quad (4)$$

$$\text{EUIM}(u) = \lambda(sui, u) \otimes \text{smt}_{su}.$$

**3.3. Potential Interest Model Construction.** PUIM is different from EUIM in which it cannot be directly obtained from the user's previous browsing content. Because of the large volume and variety of product updates and different fields, there are commodity hot spots that may have a sensational effect. Therefore, the recommended list should not only include products that are of interest to users but also products that include potential interests of users. In this regard, this paper proposes the use of collaborative filtering methods to recommend the interests of similar user groups to target users to express the potential interests of target users.

Suppose user  $u$  browses the product set  $Su = (su_1, \dots, su_l)$ ,  $\text{EUIM}_{su} = (a_{u1}, \dots, a_{ul})$ , user  $v$  reads the product set  $Sv = (sv_1, \dots, sv_k)$ ,  $Su$  and  $Sv$  are all subsets of  $S$ ,  $\text{EUIM}_{sv} = (a_{v1}, \dots, a_{vk})$ , then the behaviour similarity of users  $u$  and  $v$  is shown in formula (5). Between the formulas are operations between matrices.

$$\text{similar}_a(u, v) = \frac{(1/e^{1+U(i)})}{(Su + Sv)^{1/2}} \in [0, 1]. \quad (5)$$

Among them,  $U(i)$  represents the set of users who have viewed the product  $s_i$ . The content similarity of user's  $u$  and  $v$  is shown in the following formula:

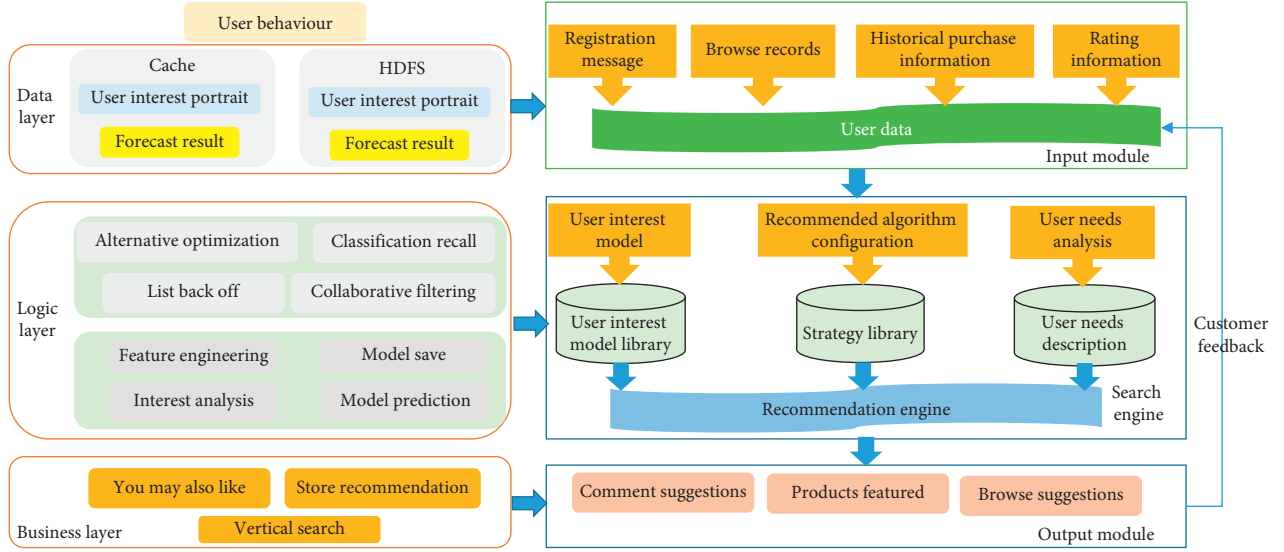


FIGURE 1: Structure diagram of the personalized recommendation system.

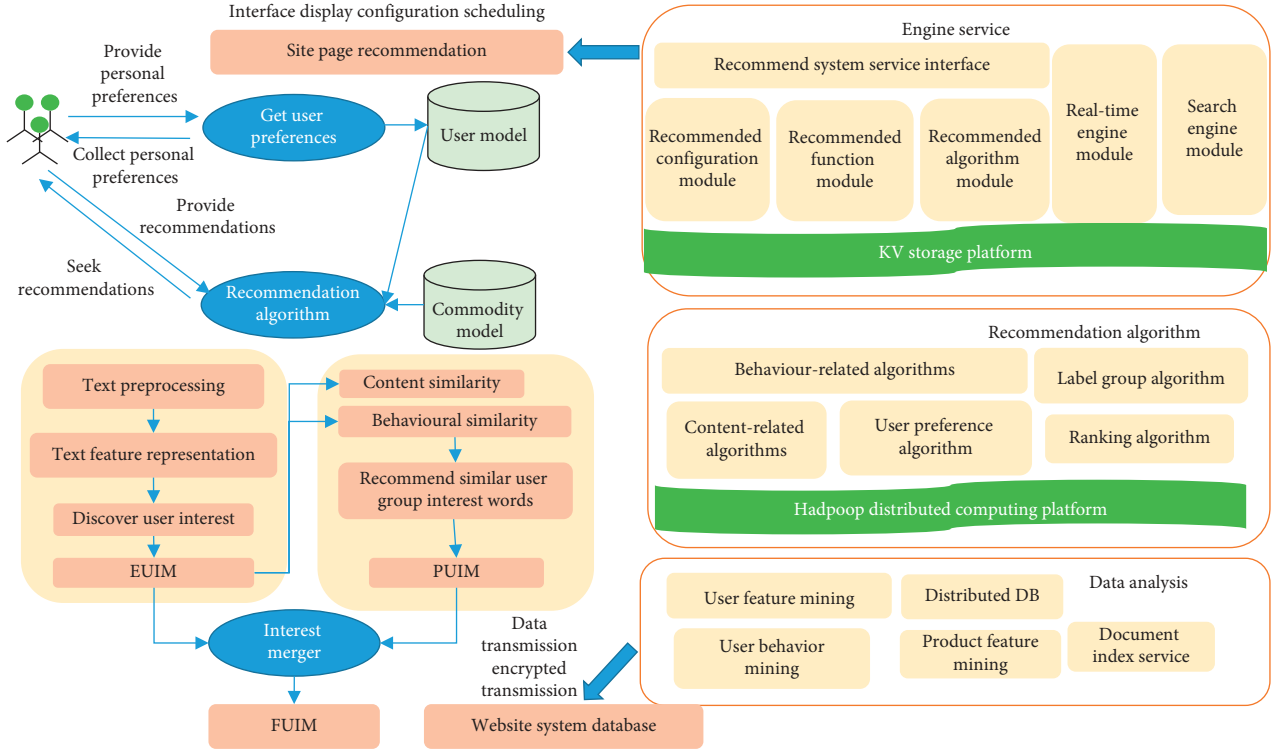


FIGURE 2: Fusion model construction process.

$$\text{similar}_c(u, v) = \frac{(1/e^{1+U(i)})}{(\text{EUIM}_{su} + \text{EUIM}_{sv})^{1/2}} \in [0, 1]. \quad (6)$$

Combining equations (5) and (6), this paper proposes a formula for calculating the hybrid similarity as shown in the following equation:

$$\text{similar}(u, v) = \alpha * \text{similar}_a(u, v) + (1 - \alpha)\text{similar}_c(u, v). \quad (7)$$

Among them, the coefficient  $\alpha$  is a weighting factor determined by experiments, which is a similarity ratio parameter, and its value range is  $[0, 1]$ . When  $\alpha = 0$ , the similarity calculation only considers content feature data. When  $\alpha = 1$ , the similarity calculation only considers behaviour characteristic data.

The behaviour similarity and content similarity of the two user's  $u$  and  $v$  are calculated and then the weighting factor  $\alpha$  is used to combine the two similarities to obtain a mixed user similarity.

The similarity between the target user and all other users is obtained through the above formula. The  $h$  users are selected with the greatest similarity to the target user as the similar user group of the target user, and collaborative filtering is used to recommend the characteristic words that are of interest to the target user to obtain the PUIM of the target user.

Suppose the similar user group is  $u = (us_1, us_2, \dots, us_j, \dots, us_f)$ , the similarity between user  $u$  and any user  $us_i$  in the similar user group is  $\text{similar}(u, us_i)$ ,  $\text{EMIM}_{us_i}$  of  $us_i = (a_1us_{i1}, a_1us_{i2}, \dots, a_1us_{ij}, \dots, a_1us_{if})$ , equation (8) is used to calculate the weight of the feature word  $M_j$  in the PUIM of user  $u$  as follows:

$$a_{us_i}^{M_j} = \frac{\text{similar}(u, us_{ij})}{\sum_{j=1}^n \text{similar}(u, us_{ij})} * \text{EMIM}_{us_{ij}}. \quad (8)$$

Among them,  $M_j$  represents the  $j$ th feature word in the PUIM of user  $u$  and  $n$  represents the number of feature words.

**3.4. Build a Fusion Model.** After obtaining the EUIM and PUIM of the target user, the weights of the feature words of the two interest models are combined to obtain the FUIM of the target user. Then, the similarity between the main feature word weight vector of the candidate product and FUIM is calculated.

Clustering operations are performed on the above similarities to form clusters and determine cluster centres. The algorithm flow chart is shown in Figure 3.

It can be clearly seen from Figure 3 that the new fusion algorithm is based on the user-feature preference matrix for clustering search. The projection characteristics derived from the fusion model, such as project characteristics and user-project rating, are intermediate results. These results will be further used in the clustering algorithm. The farthest distance principle represents the L2 distance. User preference means that when user likes product, the score for product will be higher. Then, when looking for the nearest neighbour users, the user who also has a higher score for product is given priority to become the nearest neighbour.

## 4. Results and Analysis

**4.1. Experimental Data.** The data set provided by Movie Lens website is used as the experimental data of the improved algorithm proposed in this paper for serial analysis and comparison. The Group lens research team has published three data sets of different scales, which are as follows:

- (1) The first data set includes one hundred thousand ratings of 1682 movies from 943 users
- (2) The second data set includes 1 million ratings of 3900 movies from 6040 users
- (3) The third data set includes 100000 label records and 10 million ratings of 10681 movies from 71567 users

The data set is collected by the Movie Lens website, in which each user has scored at least 20 movies he has watched, and the score is between one and five. The higher the score, the more the user prefers the movie. The data set

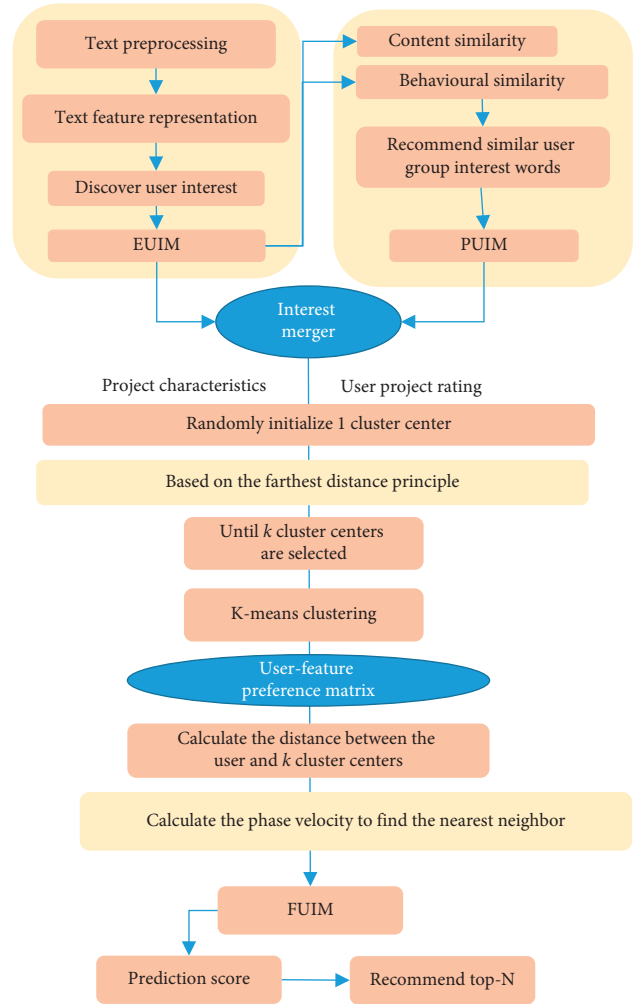


FIGURE 3: Flow chart of the recommended method.

mainly includes three data tables, namely: rating data table, user data table, and movie data table, and the composition of each table is as follows. For the Movie Lens datasets, we use the entire dataset for testing.

This paper randomly selects 11560 scores of 1682 movies from 100 users as the data basis of the experiment and randomly hides about 10% of the scores in the data set to form a test set and the remaining about 90%. The scoring data are used as the training set. According to the needs of the test, 50 users were randomly selected from the training set three times to obtain four training sets with different sparsity. Based on this, the score of each user's hidden movie was predicted. The specific distribution of the experimental data set is shown in Table 1 and Figure 4. This article selects precision, recall, and hybrid similarity for evaluation. The calculation formulas of precision and recall are as follows:

$$\text{precision} = \frac{TP}{TP + FP}, \quad (9)$$

$$\text{recall} = \frac{TP}{TP + FN}.$$

TABLE 1: Experimental data set.

Number	User number	Number of movies	Number of ratings	Sparsity
Total sample	100	1682	11560	0.931
Training set 1	100	1265	10405	0.917
Training set 2	50	1098	5354	0.902
Training set 3	50	1075	5051	0.906
Training set 4	50	1265	10405	0.835
Test set	97	567	1155	0.979

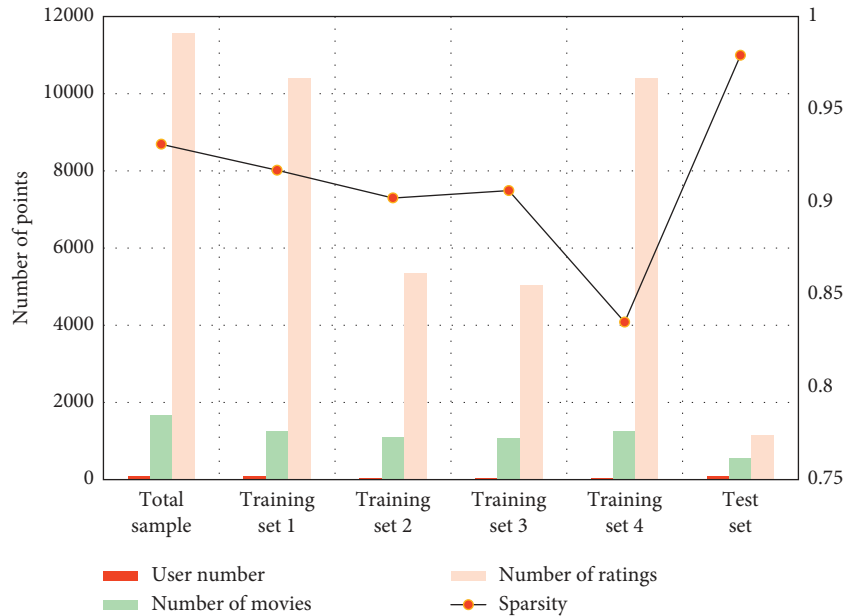


FIGURE 4: Score distribution.

Among them, TP is true positive, FP is false positive, and FN is false negative. The calculation formula of hybrid similarity will be explained in the following chapters.

**4.2. Existing Interest Model with Time Weight.** The user interest model is constructed through the traditional content-based recommendation method, and the similarity between the obtained user interest model and the candidate product is calculated to obtain the recommendation list. The recommendation list is compared with the user's actual browsing records in the test data, and the precision and recall that use traditional user interest model construction methods to generate recommendation results are obtained. Then, the method proposed in this paper is used to build an existing interest model, calculate the similarity with the candidate products, and get a recommendation list. The recommendation list is compared with the user's actual browsing records in the test data to get the precision and recall of the proposed method. Figure 5 shows the traditional user interest model and the time-weighted existing interest proposed in this paper when the number of recommended products is 5, 10, 15, 20, 25, 30, 35, and 40, respectively. Precision and recall of the recommended results are obtained by the model.

It can be seen from Figure 5 that the existing interest model with time weight proposed in this paper is compared with the user interest model constructed by the traditional content-based recommendation method to directly generate recommendation results. Recommendations are generated by the existing interest model with time weight. The results in precision and recall indicators are better than the recommendation results generated by the user interest model obtained by direct weighted average in the traditional content-based recommendation method. This proves the validity of the existing interest model with time weight.

**4.3. Mixed Similarity.** The traditional collaborative filtering method only uses the user's behaviour similarity to find similar user groups and then recommends similar user groups to the target users to browse. However, the target user has not browsed the product. This paper proposes a hybrid similarity representation based on the original behaviour similarity. The content similarity between users is compared using the user's existing interest model. By mixing the behaviour similarity and content similarity, the hybrid similarity calculation is obtained.

For recommendations using collaborative filtering, the selection of the number of similar users is very important. If there are too few similar users, the resulting

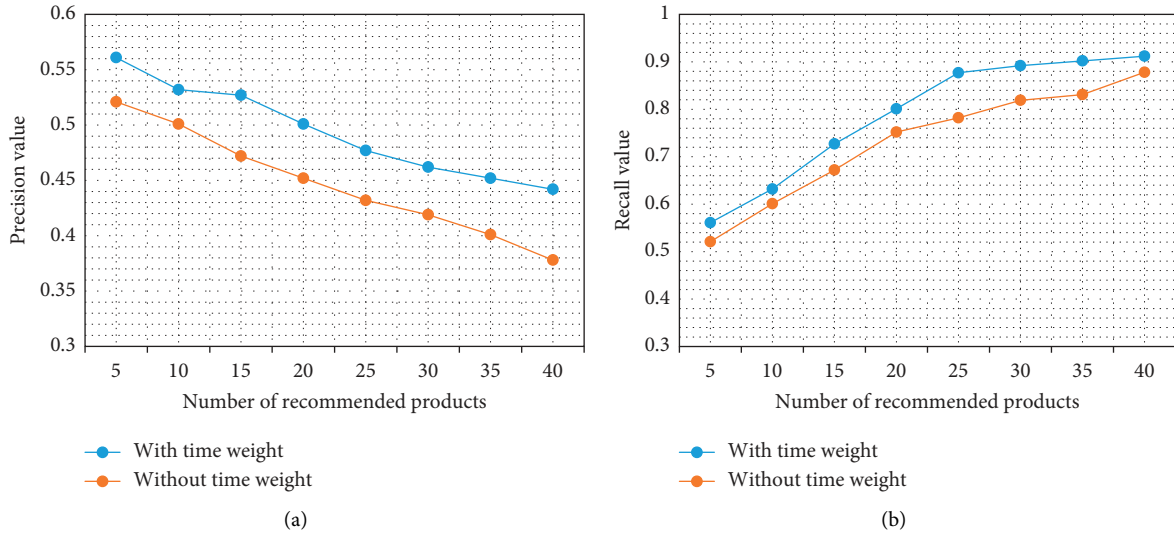


FIGURE 5: Result analysis of the recommendation algorithm with time weight and without time weight.

recommendation results are easily affected by the personal preferences of similar users. If there are too many similar users, many users with very small similarity to the target user are also included in the similar user group, which will interfere with the calculated user interest. Therefore, it is first necessary to find the optimal number of similar users through experiments. At the same time, when using mixed similarity, the value of the mixed parameter  $\alpha$  also needs to be determined through experiments. In the experiment, first  $\alpha$  is fixed to 0, 0.5, and 1, respectively, and the initial value of the number of similar users is set to 10 to generate a recommendation result and calculate the F-measure of the recommendation result. Then, with an increment by 10 people in turn, the F-measure of the recommended result is calculated again and so on to find the number of similar users when the F-measure obtains the extreme value. Figure 6 shows the F-measure of recommendation results when the number of product recommendations per user is 15,  $\alpha$  is equal to 0, 0.5, and 1, and the number of similar users is 10, 20, 30, 40, 50, 60, 70, and 80.

Through the comparison of the experimental results in F-measure in Figure 6, 60 is taken as the best number of similar users. Then, the optimal number of similar users is fixed, and  $\alpha$  is set as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9, respectively, to calculate the F-measure of the recommendation result. Table 2 shows the F-measure of recommendation results with 60 similar users and different  $\alpha$  values.

Through the comparison of the experimental results of F-measure in Table 2, it can be seen that the optimal  $\alpha$  is 0.7.

After determining the optimal number of similar users and the value of the mixed parameter  $\alpha$ , the behaviour similarity used in the traditional collaborative filtering method is directly used to find similar user groups, and products that similar users browse but have not browsed by themselves are recommended as the recommendation results to the target users. Then, the hybrid similarity proposed in this paper is used to find similar user groups and recommend products that similar users browse but have not

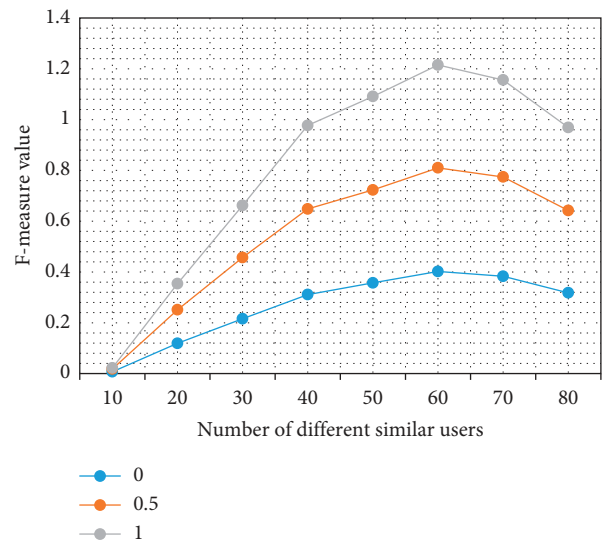


FIGURE 6: F-measure values under different  $\alpha$  values and similar users.

browsed by themselves as recommendation results to target users and get the precision and recall of the recommendation results. Figure 7 shows that when the number of recommended products is 5, 10, 15, 20, 25, 30, 35, and 40, the behavioural similarity and mixed similarity of similar user groups are directly sent to the target recommend products that similar users browse but have not browsed themselves as precision and recall as the recommendation results.

It can be seen from Figure 7 that the hybrid similarity proposed in this paper and the behavioural similarity of the traditional collaborative filtering method are used to find similar user groups, and the products that similar users browse but have not browsed themselves are directly recommended to the target as the recommendation results. From the perspective of accuracy and recall rate indicators, the hybrid similarity proposed in this paper is better than the

TABLE 2: F-measure value when the number of similar users is 60.

$\alpha$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F-measure	0.511	0.513	0.517	0.521	0.573	0.601	0.623	0.532	0.462

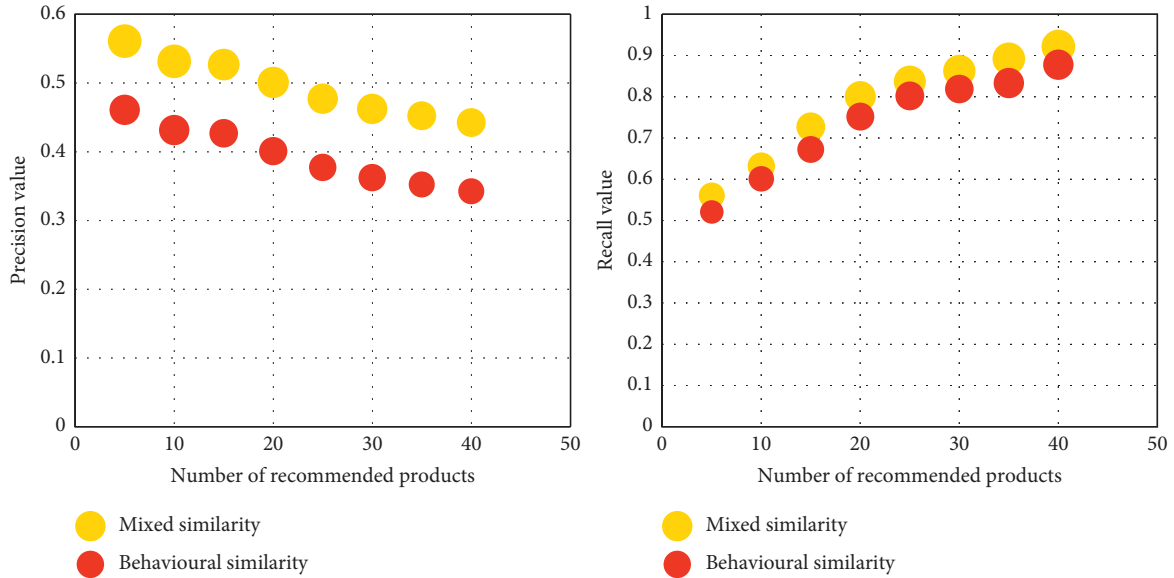


FIGURE 7: Analysis of the results of the behavioural similarity of the traditional collaborative filtering algorithm and the mixed similarity of the algorithm in this paper.

recommendation results of behaviour similarity used in the traditional collaborative filtering method. This proves the effectiveness of the hybrid similarity calculation.

**4.4. Fusion Algorithm.** The existing interest model and the potential interest model are fused to obtain a fusion interest model, and the similarity between the fusion interest model and the candidate product is calculated to obtain a recommendation list. The recommendation list is compared with the user's actual browsing records in the test data, and the precision and recall of the fusion method are obtained. Figure 8 shows the precision and recall of the recommended results by comparing the fusion method proposed in this paper with literature [10], literature [12], literature [19], and literature [21] when the number of recommended products is different.

It can be seen from Figure 8 that the fusion method and comparison method proposed in this paper are used to generate the recommendation list, and the recommended results obtained are on the precision and recall indicators. The fusion method and comparison recommendation method proposed in this paper have better results.

Finally, this paper uses literature [26], literature [27], and literature [28] as the baseline to compare with the fusion method proposed in this paper. The F-measure and

diversity of the four methods are compared, respectively, to illustrate the effectiveness of the method proposed in this paper. Figures 9 and 10 show the F-measure and diversity of the recommended results obtained by the fusion method proposed in this paper and the other three methods when the number of recommended products is different.

Figure 9 shows the F-measure of the recommended results under different methods. It can be seen that the method proposed in this paper has a significant improvement in the recommendation performance of the method in literature [27] and literature [28], which is comparable to literature [26]. Figure 10 shows the diversity of recommendation results under different methods. It can be seen that the method proposed in this paper is basically equivalent to the diversity of literature [26] and literature [28] and has a significant improvement in diversity than content-based recommendations. The method proposed in this paper finally uses the constructed fusion interest model to perform similarity matching with candidate products to generate recommendation results, and there is no cold start problem. Therefore, the actual recommendation performance of the method proposed in this paper is better than the traditional collaborative filtering recommendation, based on content recommendation and hybrid recommendation methods.



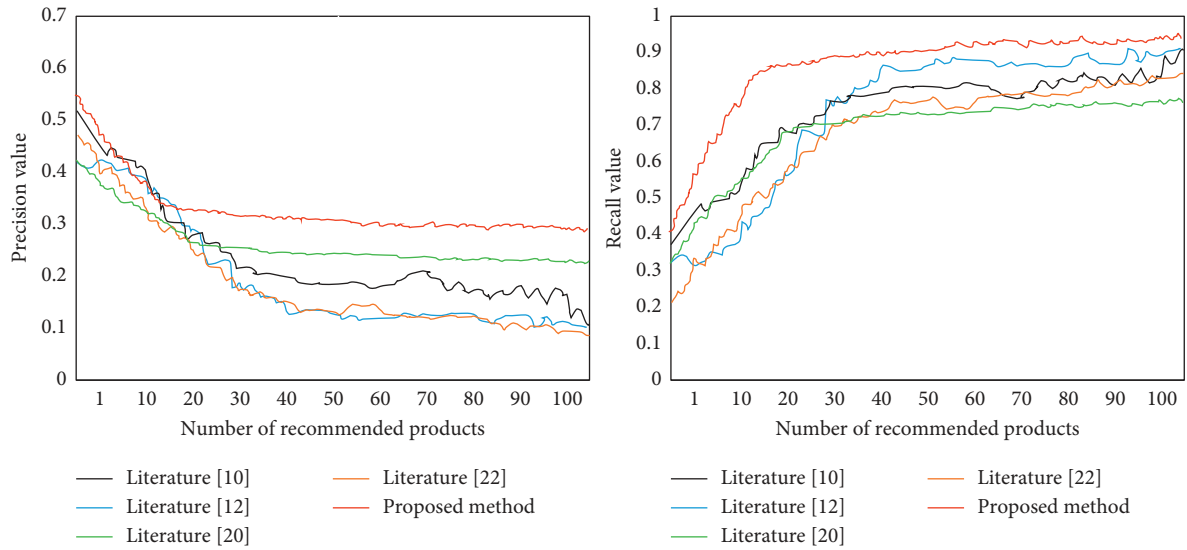


FIGURE 8: Analysis of the results of different algorithms and this algorithm.

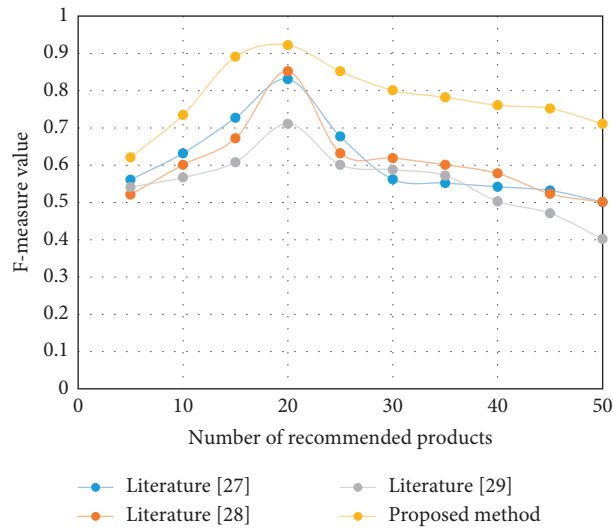


FIGURE 9: The F-measure value of this algorithm and other algorithms.

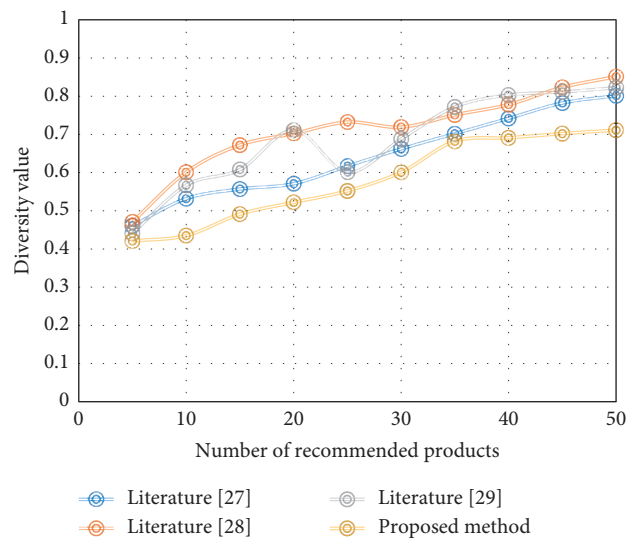


FIGURE 10: Diversity value of this algorithm and other algorithms.

## 5. Conclusion

The rapid popularity of the Internet has enabled online marketing to integrate into the lives of modern people, greatly changing the way users shop in the past and providing users with the convenience of shopping without going out. However, with the continuous expansion of the scale of e-commerce, its structure is becoming more and more complex, users are not familiar with the massive amount of product information, and merchants have lost contact with users. The wide application of the online marketing recommendation system has alleviated many problems such as “information overload” and “information trek” and enabled users to have more and better online shopping experience. It has become indispensable to help e-commerce successfully implement online marketing. At the same time, various types of online marketing recommendation systems are also facing many challenges, such as new user issues based on content filtering, data sparseness of collaborative filtering, and cold start issues. To solve the above problems, this paper proposes a fusion recommendation method based on content and collaborative filtering. This method improves the traditional content-based method to obtain the user’s existing interest and obtains the user’s potential interest through collaborative filtering of feature words. In addition, the user’s existing interest and potential interest are merged to obtain a fused user interest model. The fusion model is used to calculate the similarity of the candidate marketing content and recommend content that may be of interest to different users. Experiments show that the method proposed in this paper achieves better results than traditional content-based methods in terms of accuracy, recall, and diversity, which shows the effectiveness of this method.

## Data Availability

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

## Consent

Informed consent was obtained from all individual participants included in the study references.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Enterprise Science and Technology Commissioner Project of Tianjin under grant no. 20YDTPJC00550 and the Science and Technology Cultivation Project of TSGUAS under grant no. ZDKT2019-006.

## References

- [1] Z. Cui, X. Xu, X. U. E. Fei et al., “Personalized recommendation system based on collaborative filtering for IoT

- scenarios,” *IEEE Transactions on Services Computing*, vol. 13, no. 4, pp. 685–695, 2020.
- [2] J. Xiao, M. Wang, B. Jiang, and J. Li, “A personalized recommendation system with combinational algorithm for online learning,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 3, pp. 667–677, 2018.
- [3] S. Alian, J. Li, and V. Pandey, “A personalized recommendation system to support diabetes self-management for american indians,” *IEEE Access*, vol. 6, pp. 73041–73051, 2018.
- [4] H. Gao, Y. Xu, Y. Yin et al., “Context-aware QoS prediction with neural collaborative filtering for internet-of-things services,” *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4532–4542, 2019.
- [5] D. Wang, Y. Liang, D. Xu, X. Feng, and R. Guan, “A content-based recommender system for computer science publications,” *Knowledge-Based Systems*, vol. 157, pp. 1–9, 2018.
- [6] L. Jiang, Y. Cheng, L. Yang, J. Li, H. Yan, and X. Wang, “A trust-based collaborative filtering algorithm for E-commerce recommendation system,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 8, pp. 3023–3034, 2019.
- [7] F. Ortega, B. Zhu, J. Bobadilla, and A. Hernando, “CF4J: collaborative filtering for java,” *Knowledge-Based Systems*, vol. 152, pp. 94–99, 2018.
- [8] R. Xiong, J. Wang, N. Zhang, and Y. Ma, “Deep hybrid collaborative filtering for web service recommendation,” *Expert Systems with Applications*, vol. 110, pp. 191–205, 2018.
- [9] V. Subramaniaswamy, R. Logesh, M. Chandrashekar et al., “A personalised movie recommendation system based on collaborative filtering,” *International Journal of High Performance Computing and Networking*, vol. 10, no. 1-2, pp. 54–63, 2017.
- [10] M. K. Najafabadi, A. Mohamed, and C. W. Onn, “An impact of time and item influencer in collaborative filtering recommendations using graph-based model,” *Information Processing & Management*, vol. 56, no. 3, pp. 526–540, 2019.
- [11] C. Zhang, M. Yang, J. Lv et al., “An improved hybrid collaborative filtering algorithm based on tags and time factor,” *Big Data Mining and Analytics*, vol. 1, no. 2, pp. 128–136, 2018.
- [12] Y. Yun, D. Hooshyar, J. Jo, and H. Lim, “Developing a hybrid collaborative filtering recommendation system with opinion mining on purchase review,” *Journal of Information Science*, vol. 44, no. 3, pp. 331–344, 2018.
- [13] M. Kastner, R. Sayal, D. Oliver et al., “Sustainability and scalability of a volunteer-based primary care intervention (health TAPESTRY): a mixed-methods analysis,” *BMC Health Services Research*, vol. 17, no. 1, pp. 1–21, 2017.
- [14] M. Goyani and N. Chaurasiya, “A review of movie recommendation system: limitations, survey and challenges,” *ELCVIA: Electronic Letters on Computer Vision and Image Analysis*, vol. 19, no. 3, pp. 0018–0037, 2020.
- [15] J. B. Li, S. Y. Lin, Y. H. Hsu, and Y. C. Huang, “An empirical study of alternating least squares collaborative filtering recommendation for movielens on apache hadoop and spark,” *International Journal of Grid and Utility Computing*, vol. 11, no. 5, pp. 674–682, 2020.
- [16] Z. Tan and L. He, “An efficient similarity measure for user-based collaborative filtering recommender systems inspired by the physical resonance principle,” *IEEE Access*, vol. 5, pp. 27211–27228, 2017.
- [17] Z. Liu, L. Wang, X. Li, and S. Pang, “A multi-attribute personalized recommendation method for manufacturing service composition with combining collaborative filtering and genetic algorithm,” *Journal of Manufacturing Systems*, vol. 58, pp. 348–364, 2021.

- [18] F. Y. Liu, X. Q. Gao, and Z. Zhang, "Improved bayesian probabilistic model based recommender system," *Computer Science*, vol. 44, no. 5, pp. 285–289, 2017.
- [19] L. Tang, Z. Duan, and Y. Zhao, "Toward using social media to support ridesharing services: challenges and opportunities," *Transportation Planning and Technology*, vol. 42, no. 4, pp. 355–379, 2019.
- [20] W. Yue, Z. Wang, W. Liu, B. Tian, S. Lauria, and X. Liu, "An optimally weighted user-and item-based collaborative filtering approach to predicting baseline data for friedreich's ataxia patients," *Neurocomputing*, vol. 419, pp. 287–294, 2021.
- [21] I. D. Borlea, R. E. Precup, A. B. Borlea, and D. Iercan, "A unified form of fuzzy C-means and K-means algorithms and its partitional implementation," *Knowledge-Based Systems*, vol. 214, Article ID 106731, 2021.
- [22] P. Bhattacharjee and P. Mitra, "A survey of density based clustering algorithms," *Frontiers of Computer Science*, vol. 15, no. 1, pp. 1–27, 2021.
- [23] K. Huang, X. Ma, R. Song, X. Rong, and Y. Li, "Autonomous cognition development with lifelong learning: a self-organizing and reflecting cognitive network," *Neurocomputing*, vol. 421, pp. 66–83, 2021.
- [24] Z. Ren, "Research on fast identification technology of forged fingerprints based on the improved K-means algorithm," *International Journal of Biometrics*, vol. 13, no. 1, pp. 17–29, 2021.
- [25] A. K. Idrees and W. L. Al-Yaseen, "Distributed genetic algorithm for lifetime coverage optimisation in wireless sensor networks," *International Journal of Advanced Intelligence Paradigms*, vol. 18, no. 1, pp. 3–24, 2021.
- [26] C. A. Kolaja, B. Porter, T. M. Powell et al., "Multiple imputation validation study: addressing unmeasured survey data in a longitudinal design," *BMC Medical Research Methodology*, vol. 21, no. 1, pp. 1–9, 2021.
- [27] G. Wu and F. Li, "A randomized exponential canonical correlation analysis method for data analysis and dimensionality reduction," *Applied Numerical Mathematics*, vol. 164, pp. 101–124, 2021.
- [28] Richa and P. Bedi, "Trust and distrust based cross-domain recommender system," *Applied Artificial Intelligence*, vol. 35, no. 4, pp. 326–351, 2021.
- [29] Z. Zhang and Y. Liu, "A list-wise matrix factorization based POI recommendation by fusing multi-tag, social and geographical influences," *Journal of Internet Technology*, vol. 19, no. 1, pp. 127–136, 2018.
- [30] V. Yadav, R. Shukla, A. Tripathi et al., "A new approach for movie recommender system using K-means Clustering and PCA," *Journal of Scientific and Industrial Research*, vol. 80, no. 2, pp. 159–165, 2021.