ISSN: 2710-0588

Hybrid Recommendation Algorithm based on Product Popularity and User Preference

Xingcai Lu, Gengxin Sun, Sheng Bin*

College of Computer Science and Technology, Qingdao University, Qingdao, China

Abstract

In today's society, recommendation system is widely used to recommend users' preferred products and content. However, with the rapid development of the Internet, there are more and more contents on the Internet, and the algorithm complexity of recommending products to users through the recommendation system is getting higher and higher, and the recommendation algorithm is also more and more inclined to recommend popular products to users, which also leads to the decline in the novelty and diversity of the recommendation algorithm. In order to solve this problem, the study first adopts the method of inhibiting the recommendation ability of current popular products, and on this basis, introduces the influence of user rating on the recommendation ability of different products. At the same time, the product is classified, take the movie for example, can be classified as action film, comedy, children's film and so on. There is a very high probability that an adult will not like children's films, and a very high probability that a child will not like horror films. Impose a penalty on a product category that the user may not like, reducing the likelihood that the product will be recommended to the user. The proposed algorithm is compared with other excellent algorithms in this field on MovieLens movie score dataset, and the results of the experiments showed that the algorithm improved in accuracy, diversity, and novelty.

Keywords

Recommendation Algorithm; Bipartite Graph Network Structure; Resource Allocation; Material Diffusion.

1. Introduction

With the exponential growth of Internet information, recommendation systems are gradually being widely used in all walks of life. Nowadays, movie recommendation, e-commerce, music recommendation, social network and other fields are inseparable from recommendation system [1-2]. A good recommendation system can increase product sales and increase users' dependence on the application website. Currently, the widely used recommendation algorithms mainly include content-based recommendation algorithm [3], collaborative filtering algorithm [4], bipartite network-based recommendation algorithm [5]. The content-based recommendation algorithm calculates the correlation between items, and then recommends items with high similarity to the user based on the user's preferences. The content-based recommendation algorithm is easier to understand, but with the development of the Internet, multimedia information is becoming more and more complex, this algorithm is often difficult to understand complex multimedia information, thus failing to achieve ideal results. Collaborative filtering algorithms are further divided into item-based collaborative filtering algorithms [6-7] and user-based collaborative filtering algorithms [8]. The item-based collaborative filtering algorithm recommends to the user items that are highly similar to items that the user has liked before, while the user-based collaborative filtering algorithm recommends to the user products that other users like that are similar to the user's preferences. The collaborative filtering algorithm does not consider the data format, and data sparsity has a small impact on the recommendation results, but it also has problems such as cold start. The recommendation algorithm based on the bipartite network mainly constructs the data in the recommendation system into a bipartite network, and mines the relationship between users and products through the bipartite network to make recommendations. Recommendation algorithms based on bipartite networks have higher accuracy than traditional content-based recommendation algorithms and collaborative filtering algorithms. Most of these algorithms are based on the Mass Diffusion algorithm [5] and the heat conduction algorithm [9] is a series of algorithms derived from the basis. At present, the indicators to measure a recommendation algorithm mainly include ranking accuracy, diversity, novelty. This paper further improves the accuracy, diversity and novelty of the algorithm by improving the material diffusion algorithm and comparing it on real data sets.

2. Introduction to Material Diffusion Algorithm

2.1. Bipartite Network

The material diffusion algorithm is implemented based on the bipartite network [10-13]. The bipartite network is a special complex network [14-15]. It consists of two different parts of nodes [16]. Different nodes can be connected and the same Nodes cannot be connected to each other, as shown in Figure 1. Many networks in real life can be constructed into bipartite networks, such as movie-actor networks, disease-gene networks, scientist collaboration networks, etc.

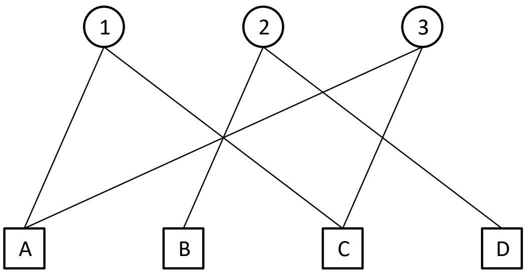

Figure 1. Bipartite network diagram

2.2. Material Diffusion Algorithm

The material diffusion algorithm first constructs a user-item bipartite network, as shown in Figure 2. There are n users and m items. The bipartite network is represented as G(U, O, E), U is the user set, O is the item set, and E denotess whether or not the users and items are connected to each other.

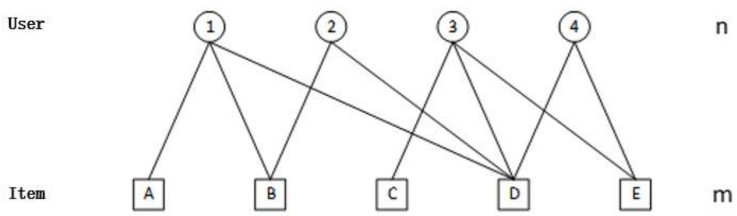

Figure 2. User-product bipartite network

Select a user node from the bipartite network. The user node sets the initial resource f_β for each project. The initial resource value f_β of the project node directly connected to the user node is set to 1. If there is no connection, f_β is set to 0, as shown in Figure 3:

ISSN: 2710-0588

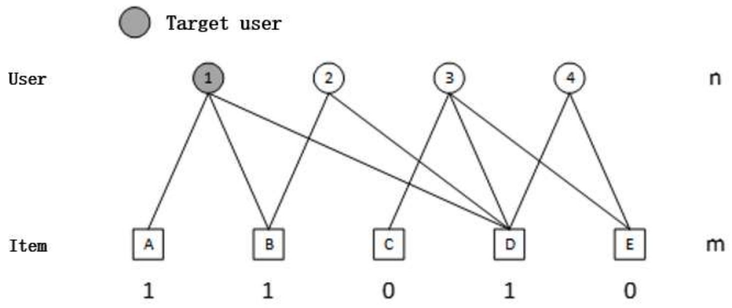

Figure 3. Initial resource assignment process diagram

Based on the number of users connected to the project, that is, the degree of the project, the initial resources are evenly distributed to each connected user. According to the resource forward conduction formula $h_i = \sum_{\beta=1}^{m} \frac{a_{i\beta} f_\beta}{k_\beta}$, calculate the total resources h_i that each user obtains from the projects he is connected to. The process is shown in Figure 4. $a_{i\beta}$ represents whether item o_β is connected to user u_i. If $a_{i\beta} = 1$, it indicatess connected, otherwise not connected. k_β represents the degree of item o_β.

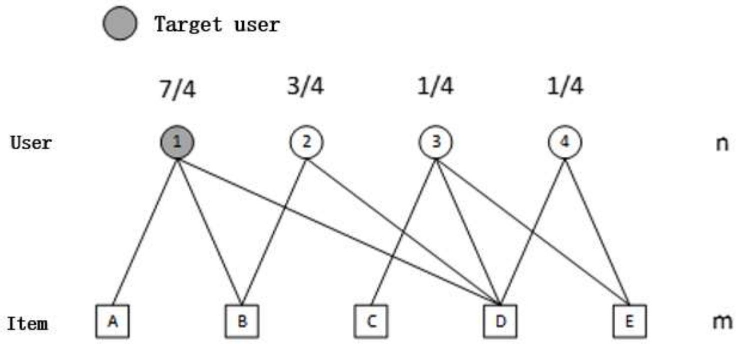

Figure 4. Resource forward transmission process diagram

Then according to the resource reverse conduction formula $f_\alpha' = \sum_{i=1}^{n} \frac{a_{i\alpha} h_i}{k_i}$, the resources are evenly allocated to each item connected to the user. Calculate the resource value f_α' of each project node, the process is shown in Figure 5. $a_{i\alpha}$ demotess whether item $a_{i\alpha}$ is connected to user u_i. If $a_{i\alpha}$=1, it means connected, otherwise not connected. k_i represents the degree of user u_i.

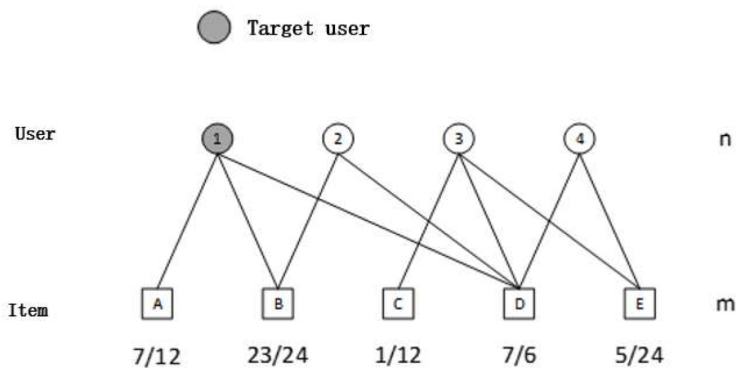

Figure 5. Resource reverse transmission process diagram

ISSN: 2710-0588

Finally, the resource values obtained by each item node are sorted from large to small. The higher the ranking, the greater the possibility of the project being recommended. It can be seen from Figure 5 that among the projects that are not connected to target user 1, the resources finally obtained by project C are $\frac{1}{12}$, and the resources finally obtained by project E are $\frac{5}{24}$. Then project E is more likely to be recommended to user 1 than project C.

Since the material diffusion algorithm sets the initial resources of each project node equally to 1, and at the same time selects the same probability for resource transfer, different nodes have the same effect when transferring resources, which makes it impossible to take advantage of the generous nodes and reduce the The novelty and accuracy of the algorithm are improved. In response to the above problems, Zhou Tao [17] introduced the concept of initial degree. By appropriately changing the initial resource value, for the target user, the initial resource of the project node connected to the user is set to $f_\beta = {k_\beta}^\lambda$, k_β is the degree of item o_β, the value of λ can adjust the recommendation ability of the project. When λ>0, the recommendation ability of large-degree nodes is improved. When λ<0, the recommendation ability of small-degree nodes is improved. Experiments show that when λ< 0, the performance of the algorithm can be further improved. When the project node transfers resources to the user node, the degree of the project node is compared with the weight of the edge. In combination, resource allocation is performed through different ratios of connected edges, which further improves the performance of the algorithm.

3. Materials and Methods

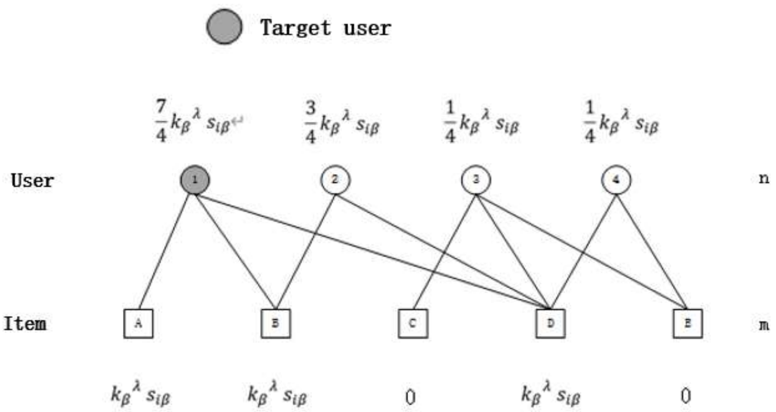

Figure 6. Improved algorithm resource forward conduction process diagram

It is assumed that the user has a score for each connected item node. The higher the score, the more satisfied the user is with the product. Otherwise, the user is dissatisfied with the product. Since the material diffusion algorithm tends to recommend currently popular products to users, but not everyone likes currently popular products, through the user's rating of the product, we can know how much the user likes the product. Therefore, in the recommendation system, we cannot ignore users' ratings of existing products. Through the above analysis, when the project node transfers resources to the user node, we combine the product node degree with the user rating. When assigning the initial resource to the product node, the initial resource assignment is $f_\beta = {k_\beta}^\lambda s_{i\beta}$, ${k_\beta}^\lambda$ is used to adjust the recommendation ability of Dadu nodes, and $s_{i\beta}$ is user u_i rating of product o_β. Combining the two can not only adjust the recommendation ability of popular products, but also add the user's love for popular products. If the higher the user's love for popular products, the recommendation algorithm will tend to recommend popular products to the user. If the user's love for popular products is not high, the recommendation ability of popular products can also be weakened. Improve the forward resource conduction formula, $h_i=\sum_{\beta=1}^{m}\frac{a_{i\beta}{k_\beta}^{\lambda}s_{i\beta}}{k_\beta}$, and calculate the sum of resources h_i that each user obtains from the projects he is connected to, The process is shown in Figure 6.

When conducting reverse resource transmission, the user conducts resource transmission to the product. At this time, we consider that users have different preferences for movie categories. For example, the probability that a child has not watched a horror movie is very high. At this time, we can think that we do not need Recommend horror movies to this child, which means reducing the likelihood of recommending horror movies. At this time we need to classify the products. For example, movies can be divided into action movies, comedy movies, science fiction movies, etc. Take the real data set MovieLens as an example. This data set classifies movies into 18 categories, including Action, Adventure, Animation, Children's, Western, etc. The ratio of the movie categories watched by target user 1 to the total number of movies watched is calculated. Statistics, as shown in Figure 7.

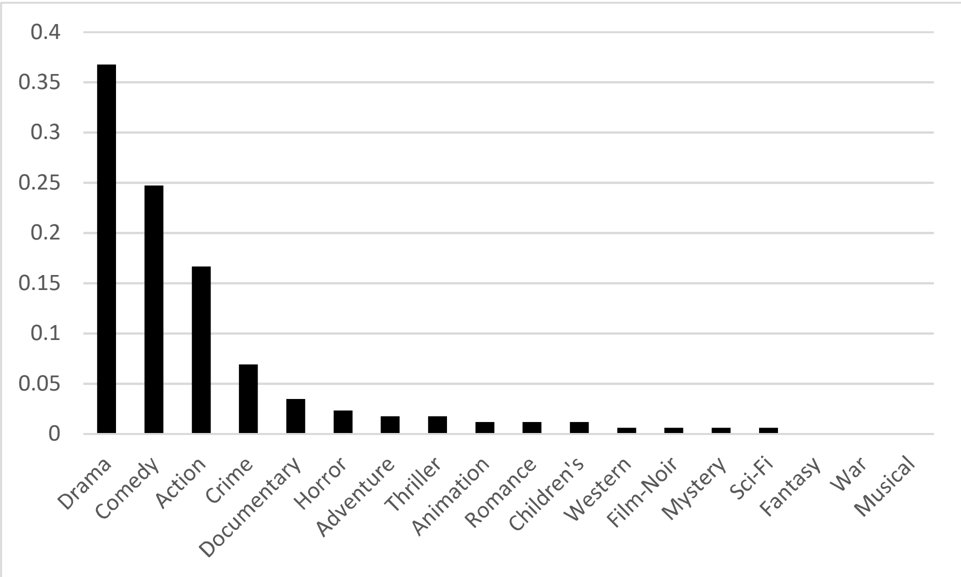

Figure 7. Target user movie classification statistical chart

As can be seen from Figure 7, the target user has not watched movies in categories such as Fantasy, War, Musical, so we can reduce the recommendation level of such movies to the target user. Since a movie may have more than one category, such as both an action movie and a comedy, as shown in Table 1:

Table 1. Classification of movies

movie title	categorization
Toy Story (1995)	Animation|Children's|Comedy
Grumpier Old Men (1995)	Comedy|Romance
Sudden Death (1995)	Action
Cutthroat Island (1995)	Action|Adventure|Romance

We need to analyze all the categories to which a movie belongs. At this time, we propose a definition for reducing the recommendation level of a certain type of movie. Assume that a certain movie belongs to c categories, and the ratio of a certain type to the total number of all movies is r, when $\sum_{i=1}^{c} r_i = 0$, we will reduce the degree of recommendation of the movie to

ISSN: 2710-0588

users. We call $\sum_{i=1}^{c} r_i$ the user's insensitivity to a certain movie. When the insensitivity is 0, we improve the reverse resource conduction formula, $f'_\alpha = \mu \sum_{i=1}^{n} \frac{a_{i\alpha} h_i}{k_i}$, (0<μ<1). μ is an adjustable parameter. When μ is smaller, the final resources obtained by the project will be smaller, which means that the movie is less likely to be recommended to users.

4. Experimental Setup

4.1. Experimental Data Set

In order to verify the effectiveness of the improved algorithm, two public data sets provided by MovieLens were selected, namely MovieLens-100k and MovieLens-1M. MovieLens-100k contains 943 users' ratings of 1,682 movies. The rating range is 1-5, where 1 represents extremely dissatisfied and 5 represents the most satisfied. The total rating records are approximately 100,000. MovieLens-1M contains ratings of 3,883 movies by 6,040 users, with a total of approximately 1,000,000 rating records. Considering the impact of users' ratings of movies on recommendation results, we select users' ratings of 3 or above to build a user-item network. Each data set is randomly selected, with 90% of the scores recorded as the training set and 10% of the scores recorded as the test set. For each algorithm and each data set, the average value is taken after ten experiments to reduce the error caused by a single experiment.

4.2. Algorithm Comparison and Evaluation Metrics

Sorting accuracy can most intuitively reflect the effect of a recommendation algorithm. Based on the recommendation results generated by the recommendation algorithm, users can easily evaluate the quality of a recommendation algorithm. In the experiment of this improved algorithm, we used the sorting accuracy to determine the optimal values of **λ** during resource forward propagation and μ during resource back propagation. Then compare the differences between different algorithms through novelty and diversity.

5. Experimental Results and Analysis

In the process of resource forward transmission, the product node degree and user ratings are combined to improve the resource forward transmission formula, and the impact of product node degree on the recommendation results is controlled through the adjustable parameter **λ**. The ranking accuracy results corresponding to the values of **λ** in the MovieLens-100k and MovieLens-1M data sets are shown in Figure 8 and Figure 9:

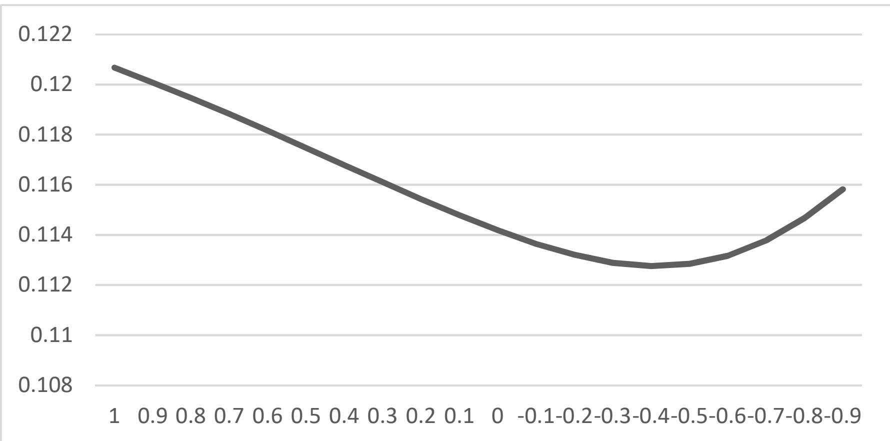

Figure 8. The impact of parameter **λ** on the algorithm under MovieLens-100k data set

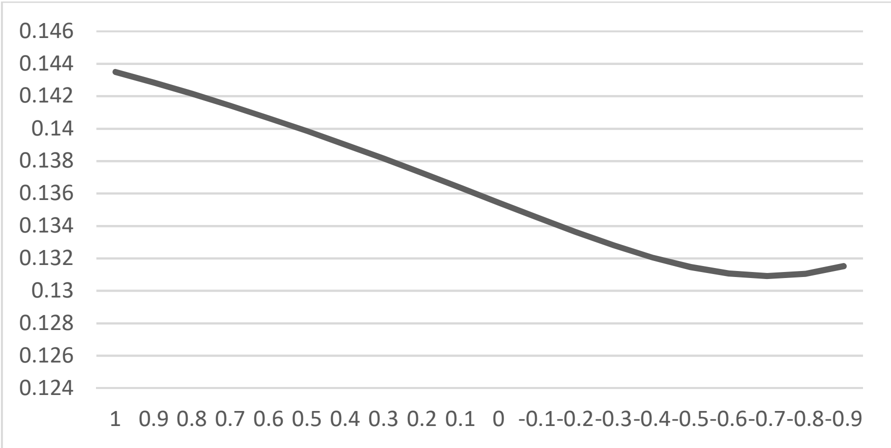

Figure 9. The impact of parameter **λ** on the algorithm under MovieLens-1M data set

In the process of resource reverse conduction, user insensitivity is introduced to improve resource reverse conduction. At the same time, the impact of product type on recommendation results is controlled through the adjustable parameter μ. First, the optimal value parameter **λ** is determined through the resource forward conduction process, and then the optimal value of μ is determined through the sorting accuracy. The ranking accuracy results corresponding to μ values in the MovieLens-100k and MovieLens-1M data set are shown in Figure 10 and Figure 11:

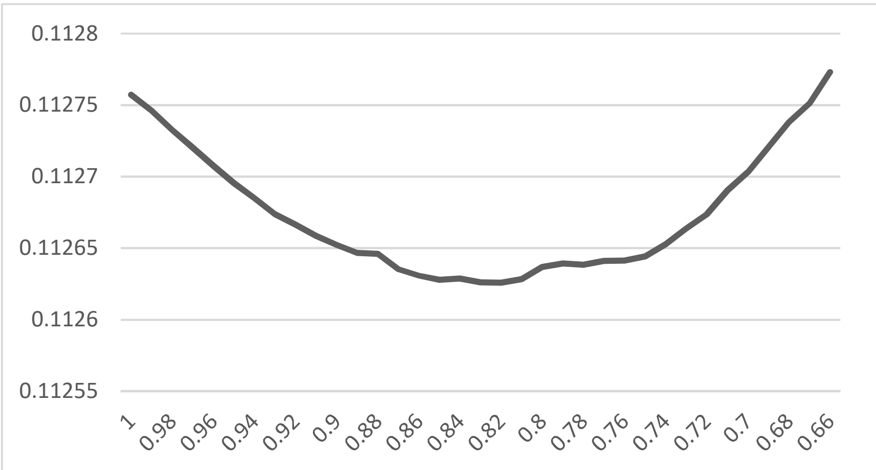

Figure 10. The impact of parameter μ on the algorithm under MovieLens-100k data set

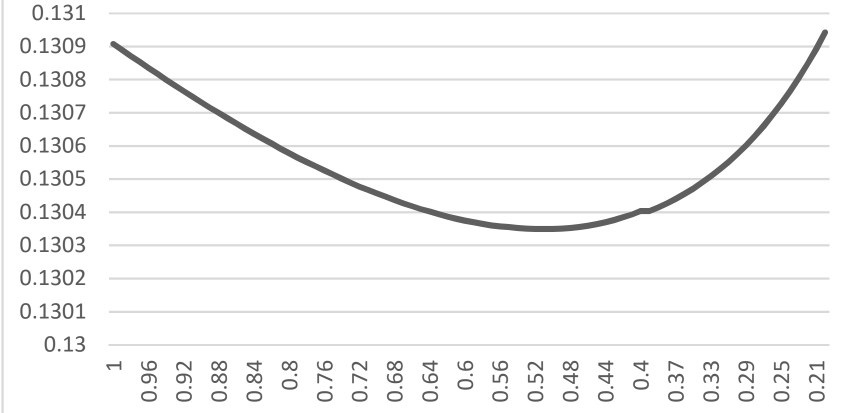

Figure 11. The impact of parameter μ on the algorithm under MovieLens-1M data set

ISSN: 2710-0588

It can be seen from Figure 8 and Figure 11 that in the MovieLens-100k data set, when **λ**=-0.4 and μ =0.82, the sorting accuracy r value reaches the best. It can be seen from Figure 9 and Figure 12 that in the MovieLens-1M data set, when **λ**=-0.7 and μ =0.51, the sorting accuracy r value reaches the best. When the sorting accuracy achieves the best value, we set the length of the recommendation list to 100, and compare it with other algorithms to obtain the sorting accuracy. The comparison results of novelty and diversity are shown in Table 2 and Table 3:

Table 2. Results of the four algorithms on the MovieLens-100k dataset

	MovieLens-100k		
	r	Novelty	Diversityh
Collaborative Filtering	0.138	1.686	0.759
Heat Conduction	0.156	0.574	0.853
Mass Diffusion	0.114	2.205	0.342
revised algorithm	0.112	2.152	0.390

Table 3. Results of the four algorithms on the MovieLens-1M dataset

	MovieLens-1M		
	r	Novelty	Diversityh
Collaborative Filtering	0.131	10.816	0.796
Heat Conduction	0.145	4.596	0.870
Mass Diffusion	0.135	14.902	0.245
revised algorithm	0.130	14.251	0.362

It can be seen from the data in Table 2 that in the MovieLens-100k data set, the improved algorithm has improved sorting accuracy, novelty, and diversity compared with the unimproved material diffusion algorithm. In terms of sorting accuracy, it is 18.8% higher than the collaborative filtering algorithm, 28.2% higher than the heat conduction algorithm, and 1.7% higher than the material diffusion algorithm.

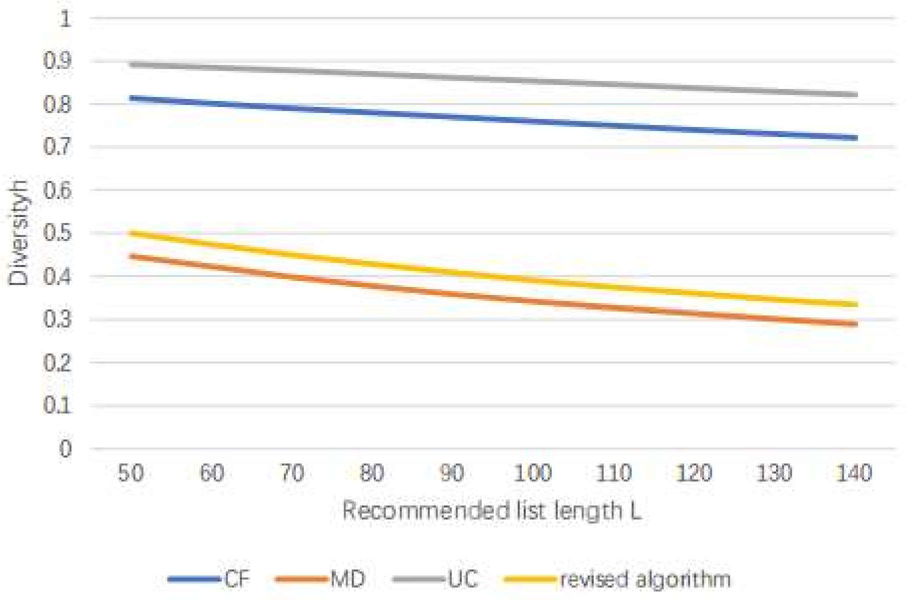

Figure 12. Comparison of diversity of different recommended lengths

ISSN: 2710-0588

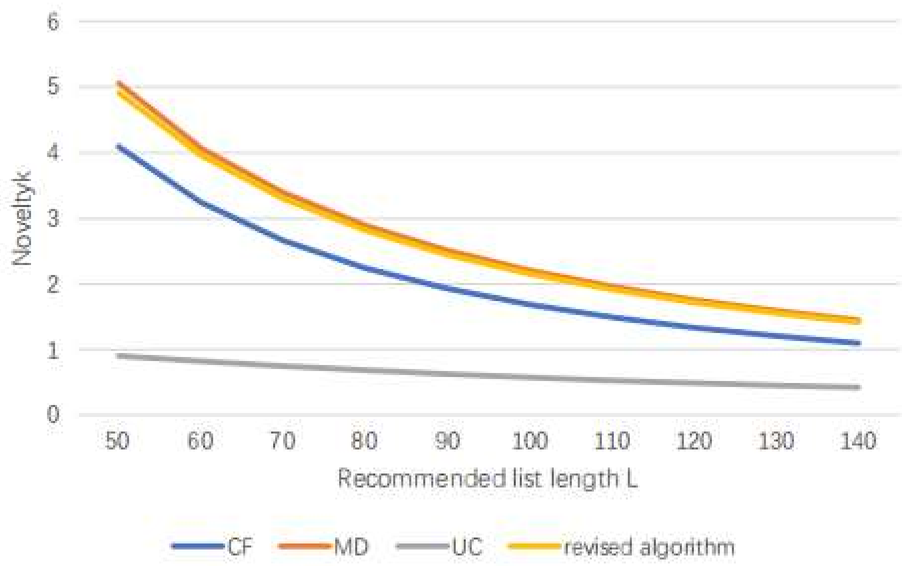

Figure 13. Comparison of novelty of different recommended lengths

It can be seen from the data in Table 3 that in the MovieLens-1M data set, the improved algorithm has improved ranking accuracy, novelty, and diversity compared with the unimproved material diffusion algorithm. In terms of sorting accuracy, compared with the collaborative filtering algorithm, it decreased by 0.7%, the specific heat conduction algorithm increased by 9.6%, and the specific material diffusion algorithm increased by 3.7%.

Figure 12 and Figure 13 show the experimental comparison results of different recommendation list lengths under the dataset MovieLens-100k for the diversity and novelty of several algorithms.

It can be seen from Figure 12 that as the length of the recommendation list increases, the diversity of the algorithm decreases and the novelty increases. But on the whole, improving the algorithm is better than improving the diversity and novelty of the previous material diffusion algorithm. When the length of the recommendation list is 100, the diversity of the improved algorithm is increased by 14.1% and the novelty is increased by 2.4% compared with the material diffusion algorithm before the improvement. However, the improved algorithm is not as good as the collaborative filtering algorithm and heat conduction algorithm in terms of novelty and diversity.

6. Conclusion

By improving the material diffusion algorithm, this paper introduces non-uniform initial resources and user ratings to forward conduction on the original basis, and combines the initial resources and user ratings. Although the recommendation ability of the generous node is reduced, if If the user's love value for popular products is large, it will further improve the recommendation ability of Dadu nodes. If the user's love value for popular products is small, it will further reduce the recommendation ability of Dadu nodes. At the same time, the concept of user insensitivity is introduced for reverse conduction. By classifying products, the recommendation ability is reduced for product categories that users may not like. Judging from the experimental results, through two improvements, not only the sorting accuracy of the algorithm has been improved, but the diversity and novelty capabilities have also been further improved.

ISSN: 2710-0588

References

[1] Adomavicius G , Tuzhilin A . Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6):734-749.

[2] Resnick P , Varian H R . Recommender systems. Commun ACM[J]. Communications of the ACM, 1997, 40(3):56–58.

[3] Lang K . Learning to filter netnews[J]. proc icml, 1994.

[4] Al-Shamri M . Power coefficient as a similarity measure for memory-based collaborative recommender systems[J]. Expert Systems with Applications, 2014, 41(13):5680–5688.

[5] T Zhou, J Ren, M Medo, et al. Bipartite network projection and personal recommendation.

[6] Sarwar B M , Karypis G , Konstan J A , et al. Item-based collaborative filtering recommendation algorithms. 2001.

[7] Deshpande M , Karypis G . Item-based top-N recommendation algorithms[J]. ACM Transactions on Information Systems (TOIS), 2004.

[8] Blei D M , Ng A Y , Jordan M I , et al. Journal of Machine Learning Research 3 (2003) 993-1022 Submitted 2/02; Published 1/03 Latent Dirichlet Allocation. 2003.

[9] Tao Z , Kuscsik Z , Liu J G , et al. Solving the apparent diversity-accuracy dilemma of recommender systems[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(10):4511-4515.

[10] Dorogovtsev S N , Goltsev A V , Mendes J . Critical phenomena in complex networks[J]. Rev.mod.phys, 2007, 80(4):1275-1335.

[11] Carsten F. Dormann,Rouven Strauss. A method for detecting modules in quantitative bipartite networks[J]. Methods in Ecology and Evolution,2014,5(1).

[12] Evans TS , Lambiotte R.Line graphs,link partitions,and overlapping communities. Phys Rev E 80(1):016105[J]. Physical Review E, 2009, 80(1 Pt 2):016105.

[13] Fu C H , Zhang Z P , Chang H , et al. A kind of collaboration–competition networks[J]. Physica A Statistical Mechanics & Its Applications, 2008, 387(5-6):1411-1420.

[14] ZHOU T. A brief review of complex networks[J].Physics, 2005, 34(1):31-36.

[15] Fang, Jinqing. Theoretical research progress in complexity of complex dynamical networks[J]. Progress in Natural Science, 2007, 17(7):841-857.

[16] Tao Z , Kuscsik Z , Liu J G , et al. Solving the apparent diversity-accuracy dilemma of recommender systems[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(10):4511-4515.

[17] Zhou T , Jiang L L , Su R Q , et al. Effect of initial configuration on network-based recommendation[J]. EPL (Europhysics Letters), 2008, 81(5):58004.