International Journal of Smart Business and Technology
Vol.8, No.2 (2020), pp.1-12
http://dx.doi.org/10.21742/ijsbt.2020.8.2.01

A Collaborative Filtering Recommendation Algorithm based on Cluster Analysis

Matteo Scarpa[1] and Felice Antonio Caserta[2]

[1,2]University of Pisa, Pisa, Italy
[1]matteo.scarpa@unifi.it, [2]antonio.caserta61@gmail.com

Abstract

As the application of personalized recommendation systems in e-commerce websites becomes more and more extensive, the research on personalized recommendation algorithms is constantly deepening. The personalized recommendation system has brought huge commercial benefits to the e-commerce field. Based on the item-based top-N collaborative filtering recommendation algorithm, this paper proposes a top-N collaborative filtering recommendation algorithm based on the K-means clustering algorithm. Using K-means to cluster according to the distance between the sample points, the similarity between users is regarded as the distance, the users are clustered into several clusters, and then the recommendation algorithm is applied in each cluster to perform recommendations. This paper introduces factors such as item time difference, popular item weight, and user common rating weight into the similarity measurement formula. The experimental results show that the recall rate of the algorithm proposed in this paper is higher than that of the traditional collaborative filtering recommendation algorithm2. 1%, proved the correctness of the proposed algorithm.

Keywords: *Cluster analysis, Similarity calculation, Collaborative filtering algorithm, Item similarity, Threshold, Recall rate*

1. Introduction

The personalized recommendation is currently one of the most effective methods to solve the problem of information overload. Unlike search engines, the recommendation system does not require users to provide clear requirements but is based on the user's historical feedback information to model the user's interest, to actively recommend information that can meet their interests and needs. As the application of personalized recommendation systems in e-commerce websites becomes more and more extensive, the research on personalized recommendation algorithms is constantly deepening. The personalized recommendation system has brought huge commercial benefits to the e-commerce field. According to VentureBeat statistics, Amazon's recommendation system provides 35% of its merchandise sales. In addition to recommending products for users on e-commerce sites, you can also recommend books, movies, news, and music for users.

The collaborative filtering recommendation algorithm is currently one of the most widely used and most successful personalized recommendation algorithms. The concept of collaborative filtering was proposed by Coldberg, Nieols, Oki, and Terry in 1992 and applied to the Tapestry system. Collaborative filtering recommendation algorithms include memory-based collaborative filtering recommendation algorithms and model-based collaborative

Article history:
Received (June 5, 2020), Review Result (July 9, 2020), Accepted (August 15, 2020)

Print ISSN: 2288-8969, eISSN: 2207-516X IJSBT

filtering recommendation algorithms. Memory-based collaborative filtering algorithms can also be divided into item-based and user-based collaborative filtering recommendation algorithms [2][3]. The user-based recommendation algorithm calculates the similarity between users to find the user group most similar to the target user according to the similarity between users, and finally recommends the target user that similar users have purchased but the target user has not purchased Items. Similar to the user-based algorithm, the item-based filtering algorithm first calculates the similarity between items, determines the most similar item group of the items that the target user has purchased according to the similarity between the items, and finally selects the most similar item group Some items are recommended to target users.

The classic collaborative filtering recommendation algorithm has problems such as cold start, data sparsity, and algorithm scalability. Many studies have partially solved the above-mentioned problems based on a collaborative filtering recommendation algorithm and improved the recommendation quality. Literature [4] uses the clustering algorithm to solve the scalability problem of the traditional collaborative filtering recommendation algorithm, and also reduces the time complexity of the algorithm. Literature [5][6] proposed a multi-clustering algorithm, which aggregates all users and items into several user-item subgroups. Each subgroup contains part of the user and item data. Experiments show that it is better than the original algorithm. The accuracy of the recommended results has been improved. Literature [7] proposed a new algorithm to solve the cold start problem of the traditional collaborative filtering recommendation algorithm. Literature [8] uses the average of user ratings as the demarcation point to get the score difference between users, and mines the user's interest in the item category and user-item preferences based on user-item ratings and item category information, and finally obtains an improved similarity calculation method, The experimental effect has been significantly improved. Literature [9] proposed a collaborative filtering recommendation algorithm that combines item attributes and preferences. Firstly, item preferences are constructed using item basic features and user characteristics; then item similarity fusion is used to improve similarity calculation accuracy.

User groups with similar interests to the target user have a higher reference value than other users. At the same time, the measurement of similarity between users is the most critical step in the user clustering algorithm. The above-mentioned documents are calculated using traditional similarity measurement methods. The similarity between users. Setting specific similarity measurement methods for different application scenarios can improve the accuracy of similarity measurement, thereby improving the quality of recommendations. This paper proposes to combine the K-means clustering algorithm with the top-N collaborative filtering recommendation algorithm. Firstly, factors such as item time difference, popular item weight, and user's common rating weight are introduced into the calculation formula of item and user similarity to improve the accuracy of similarity calculation. Secondly, users are clustered into clusters with similar interests according to the similarity between users. Finally, the recommendation algorithm is applied in the cluster class to make recommendations for users respectively. The experimental results show that the recommendation results obtained by using the top-N collaborative filtering recommendation algorithm combined with the clustering algorithm have improved the recall rate.

2. Theoretical basis

The calculation of similarity is a key step of the traditional collaborative filtering recommendation algorithm. The most commonly used methods are cosine similarity, Pearson correlation, and modified cosine similarity [10][11][12].

Assuming that there are a total of n users and m items, the user's behavior information on items is stored in a matrix $H \in R^{n \times m}$.

$$H = \begin{bmatrix} r_{1,1} & r_{1,R} & r_{1,3} & \cdots & r_{1,m-1} & r_{1,m} \\ r_{2,1} & r_{2,2} & r_{2,3} & \cdots & r_{2,m-1} & r_{2,m} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ r_{n-1,1} & r_{n-1,2} & r_{n-1,3} & \cdots & r_{n-1,m-1} & r_{n-1,m} \\ r_{n,1} & r_{n,2} & r_{n,3} & \cdots & r_{n,m-1} & r_{n,m} \end{bmatrix}$$

$U = \{u_1, u_2, u_3 \cdots, u_n\}$ represents the user set, $I = \{i_1, i_2, i_3 \cdots i_m\}$ represents the item set, and $r_{i,j}$ represents the user u_i's rating of the item i_j.

(1) Cosine similarity

When the cosine formula is used to calculate the similarity between users, each user is regarded as an m-dimensional vector, which is represented by U_i and U_j respectively, and the similarity between user u_i and user u_j is

$$sim(u_i, u_j) = cos(u_i, u_j) = \frac{u_i \cdot u_j}{\|u_i\|_2 \|u_j\|_2} \tag{1}$$

Similarly, when it is necessary to calculate the similarity between items, consider each item as an n-dimensional vector, and then use the above formula to calculate the similarity between items.

(2) Pearson correlation

Assuming that the set of items scored by user u_i and user u_j is $I_{i,j}$, the calculation formula for calculating user similarity using Pearson correlation is

$$sim(u_i, u_j) = \frac{\sum_{c \in I_{i,j}} (r_{i,c} - \overline{r_i})(r_{j,c} - \overline{r_j})}{\sqrt{\sum_{c \in I_{i,j}} (r_{i,c} - \overline{r_i})^2} \sqrt{\sum_{c \in I_{i,j}} (r_{j,c} - \overline{r_j})^2}} \tag{2}$$

$r_{i,c}$ and $r_{j,c}$ represent the ratings of item i_c by user u_i and user u_j respectively; r_i and r_j represent the average ratings of the items by users u_i and u_j, respectively.

(3) Modified cosine similarity

In the cosine similarity measurement method, the rating scale of different users is not considered. The modified cosine similarity measurement method improves the above-mentioned defects by subtracting the user's average rating of the item. Assuming that the set of items scored by users u_i and u_j is $I_{i,j}$, I_i, and I_j represent the set of items scored by users u_i and u_j, respectively, and the similarity calculation formula for users u_i and u_j is

$$sim(u_i, u_j) = \frac{\sum_{c \in I_{i,j}} (r_{i,c} - \overline{r_i})(r_{j,c} - \overline{r_j})}{\sqrt{\sum_{c \in I_i} (r_{i,c} - \overline{r_i})^2} \sqrt{\sum_{c \in I_i} (r_{j,c} - \overline{r_j})^2}} \tag{3}$$

The experimental results show that the experimental effect obtained by using Pearson correlation to calculate the similarity is better than the other two methods [12]. In the following experiments, this article will use Pearson correlation to calculate the similarity between users or items.

3. Improved collaborative filtering recommendation algorithm

There are two most commonly used recommendation methods for collaborative filtering recommendation algorithms: One is that the target user predicts and scores items that have not been evaluated, and then recommends the target user based on the predicted score. The other is the recommendation set in the form of top-N, which separately counts the interest of the target users to users who have not purchased or evaluated them, and takes the top N items to form a recommendation set. This article adopts the second recommended form.

3.1. Improved similarity calculation method

Aiming at the problems of traditional similarity calculation methods, this paper proposes an improved method for similarity calculation. In the calculation formula of the similarity between users and items, the time difference of items, the weight of popular items, and the weight of users' common rating are respectively introduced.

Literature [3] proposed that the experimental results obtained by recommending the most popular items for users are far inferior to the classic collaborative filtering recommendation algorithm because this method cannot provide users with personalized recommendations. When two users purchase only a few of the most popular items at the same time, the traditional similarity measurement method is used to calculate the similarity relationship, which will cause the similarity to be high. There is no personalized similarity between the two users, that is, they do not necessarily have a similar relationship of interest. Therefore, in the traditional user similarity calculation method, the weight of popular items is introduced.

The most popular items in matrix H are items with a higher degree. Define the degree of item i_c as $d_c = \sum_{j=1}^{n} H_{j,c}$, and the average degree of all items $d_{\text{avg}} = \frac{1}{m}\sum_{i=1}^{m} d_i$. The definition of the weight of popular items is shown in equation (4).

$$WeigPopu_{i,j} = 1 - \frac{\sum_{k \in I_{i,j}} P_k}{|I_{i,j}|} \tag{4}$$

$$P_k = \begin{cases} 1 & d_k > d_{avg} \\ 0 & d_k \leqslant d_{avg} \end{cases} \tag{5}$$

$|I_{i,j}|$ represents the number of elements in the set of items scored by users u_i and u_j. Equation (4) reflects that in the calculation of user similarity, the influence of popular item weight on user similarity should be reduced.

The traditional similarity measurement method has another problem, that is, if the number of items scored by two users is very small, the user similarity obtained according to equation (2) will be very high. To solve this problem, a common user rating weight factor is introduced. The weight of the user's common rating is defined as

$$WeigComm_{i,j} = \frac{|I_{i,j}|}{|I_i \cup I_j|} \tag{6}$$

The improved user similarity calculation method is

$$simIm(u_i, u_j) = WeigPopu_{i,j} \times WeigComm_{i,j} \times sim(u_i, u_j) \tag{7}$$

$\text{WeigPopu}_{i,j} \in [0, 1]$ is a popular item weight factor; $\text{WeigComm}_{i,j} \in [0, 1]$ is the weighting factor of the user's common rating; $\text{sim}(u_i, u_j)$ is the user similarity calculated by the formula (2).

Time information is an attribute of items. The more similar the appearance time of similar items, the higher the similarity between items, because many items have certain timeliness and popularity. Taking movies as an example, movies of the same subject matter will be released in one time period. During the New Year, comedy movies are the majority, and during Valentine's Day, romance movies are the majority. This article introduces the item's time difference factor into the calculation formula of item similarity and tries to improve the accuracy of item similarity calculation.

Assuming that the appearance times of items i_c and i_d are t_c and t_d, the greater the time difference between the appearance of the two items, the smaller the similarity between them. The impact factor of the item time difference is defined as

$$fac\,T_{c,d} = e^{-|t_c - t_d|} \tag{8}$$

Similarly, in the formula for calculating the similarity of items, the weight of users' simultaneous ratings is introduced. Suppose that the set of users who rate items i_c and i_d at the same time is $U_{c,d}$, U_c and U_d denote the set of users who rate items i_c and i_d, respectively, and the weight factor of the users' simultaneous rating is defined as

$$facComm_{c,d} = \frac{|U_{c,d}|}{|U_c \cup U_d|} \tag{9}$$

Finally, the improved formula for calculating the similarity of items is

$$\mathrm{simIm}(i_c, i_d) = \mathrm{facComm}_{c,d} \times \mathrm{fac}\,T_{c,d} \times \mathrm{sim}(i_c, i_d) \tag{10}$$

$\mathrm{facComm}_{c,d} \in [0, 1]$ is the weight of users' ratings at the same time; $\mathrm{fac}\,T_{c,d} \in (0, 1)$ is the item time difference factor; $\mathrm{sim}(i_c, i_d)$ is the similarity of items calculated by the formula (2).

3.2. Algorithm introduction

When the traditional collaborative filtering recommendation algorithm recommends for the target user, the purchase records of all users serve as a reference for the target user. However, there are similarities between users, and users who are more similar to the target user have a more valuable reference role than other users. Therefore, this paper proposes a top-N collaborative filtering recommendation algorithm based on the K-means clustering algorithm. The basic idea of the algorithm is to first use the K-means clustering algorithm to cluster according to the distance between the sample points, regard the similarity between users as the distance, and cluster the users in the item matrix H into several clusters. When it is necessary to make a recommendation for a certain target user, first calculate the distance between the target user and each cluster center to determine the cluster category to which the target user belongs. Finally, the top-N recommendation algorithm is applied in the cluster to make recommendations for the target users.

One problem that needs to be noted is that when a new user joins, the collaborative filtering recommendation algorithm cannot provide recommendations for him. This is the cold start problem. Given this situation, the most direct solution is adopted to recommend the current N most popular items for new users. The flow of the algorithm is as follows:

Input: user rating matrix H, cluster number p, and most similar item collection size K.

Output: Top-N recommendation results for target users.

(1) Calculate the similarity between users according to formula (7), and apply the K-means clustering algorithm to cluster all users into p categories. The similarity between users represents the distance between the points of K-means. The greater the similarity between two

users, the smaller the distance between the two points. Use U to denote the user set, and use the clustering algorithm to divide the user set U into $U_1, U_2, \cdots, U_p,\ and\ U_i \cap U_j = \emptyset, 1 \leqslant i, j \leqslant p, U_1 \cup U_2 \cup \cdots \cup U_p = U_0$

(2) Use formula (2) to calculate the similarity between the target white and the cluster center. Assuming that the target user has been determined $u_i \in U_j(1 \leqslant i \leqslant n, 1 \leqslant j \leqslant p)$, The information matrix between the white and the objects in the set U_j is denoted as $H_i \in \mathrm{R}^{n_i \times m}$. Among them, n_i is the number of users in the set U_i.

(3) Define the set of items that user u_i has purchased as R, then the set of items that user u_ii has not purchased is S. According to formula (10), calculate the similarity between each item in the matrix H_i, and record it as the item similarity matrix $W \in \mathrm{R}^{m \times m}$. $W_{i,j} \in [0, 1], 1 \leqslant i, j \leqslant m$. The matrix W is a symmetric matrix, and the values of the elements on the diagonal are all 1.

(4) Arrange each row of matrix W in order from largest to smallest, and the first one is yourself, because the similarity between the same items is, get the sorted item similarity matrix $\overline{W} \in \mathrm{R}^{m \times m}$ 。

(5) Determine the user u_i's favorite degree $l_{i,j}$ for each item $j \in S$. The calculation method is shown in equation (11).

$$l_{i,j} = \sum_{t \in R \cap K(j)} W_{j,t} \cdot H_{i,t} \tag{11}$$

K(j) represents the K most similar item sets of the item i_j, that is, the item set corresponding to the second column to the K 1th column in the row of item i_j in the matrix W.

Sort the items in the set S according to the user u_i's degree of interest, and select the top N item sets as the final recommendation set. It should be noted that when user u_i is a new user because there is no historical purchase information of u_i in matrix H, it is impossible to provide recommendations for him. Here, recommend the most popular N items for it.

4. Experimental results and analysis

In this chapter, the top-N collaborative filtering recommendation algorithm based on the K-means clustering algorithm is compared with the classic top-N collaborative filtering recommendation algorithm, and the experimental results obtained are analyzed.

4.1. Experimental design

This article uses recall rate, accuracy rate, and coverage rate [3] to evaluate the experimental results. Recommend N items to the target user u_i, denoted as $N(i)$。 Take N as 10 in the experiment. Let the target user u_i's favorite item set on the test set be $T(i)$, the recall rate and accuracy rate are calculated as follows:

$$recall = \frac{\sum_{i \in U} |N(i) \cap T(i)|}{\sum_{i \in U} |T(i)|} \tag{12}$$

$$precision = \frac{\sum_{i \in U} |N(i) \cap T(i)|}{\sum_{i \in U} |N(i)|} \tag{13}$$

Coverage is another evaluation criterion that needs to be considered. It is proposed in response to the long-tail problem [13], that is, some popular items are always recommended, while some unpopular items have not been paid attention to by users. Coverage rate is an evaluation index that reflects the algorithm's processing of long-tail problems. The calculation formula for defining coverage rate is as follows:

$$coverage = \frac{|\cup_{i \in U} N(i)|}{|I|} \tag{14}$$

In the algorithm, three parameters have an impact on the experimental results. The first is the top-N recommendation algorithm. After calculating the similarity between each item, the number of items in the most similar item set K is determined for each item. In the experiment, K is the number of user clusters p. This paper clusters all users into 2, 3, 4, 5, and 6 clusters respectively. The last parameter is the number of steps in the K-means algorithm, which are set to 5, 10, 15, 20, ..., 50 steps. In the experiment, we constantly test to find the best parameter settings for the algorithm and compare it with the top-N collaborative filtering recommendation algorithm.

The experiment uses movie data collected by the GroupLens group of the University of Minnesota, which has three data sets of data size: 100 KB, 1 MB, and 10 MB. Take a 100 KB data set as an example. In this data set, there are 943 users' rating sets for 1682 movies. The rating interval is 15, and there are a total of 100,000 rating items.

The data used in this algorithm is a data set of 1 MB in size. This data set has 6040 users' ratings of 3,952 movies, and the sparsity of this data set is about 94%. In this article, 80% of the data is used as the training set, and 20% of the data is used as the test set.

4.2. Experimental results

First, the recall rate and accuracy rate are used as the evaluation criteria to test the value of the parameter K that makes the experimental result the best. Nine different K values are used to test the top-N collaborative filtering recommendation algorithm, and the experiment is repeated five times, and the average of the recall rate and accuracy rate obtained from these five experiments is taken as the final result. The experimental results obtained are shown in [Figure 1] and [Figure 2].

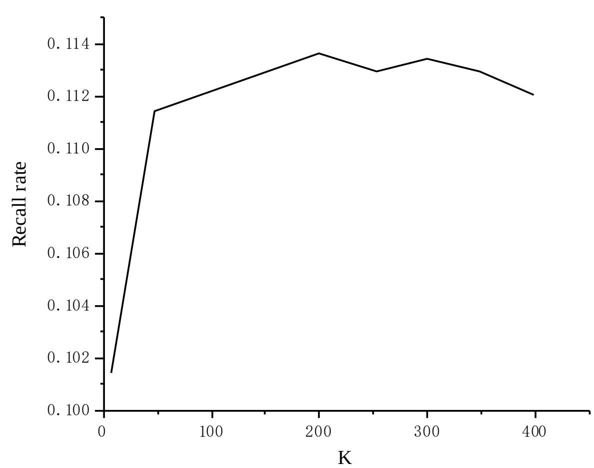

Figure 1. The recall rate of recommended results under different K values

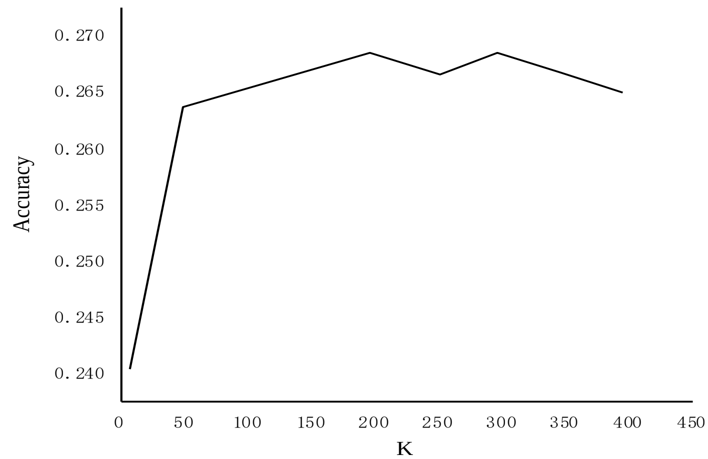

Figure 2. The accuracy of the recommended results under different K values

It can be seen from [Figure 2] that near the K value of 200, the recall rate and accuracy of the classic item-based top-N collaborative filtering recommendation algorithm reach the highest. Therefore, the top-N collaborative filtering recommendation algorithm based on K-means will also take the K value of 200, that is, the number of most similar item sets for each item in the collaborative filtering recommendation algorithm is 200. From equations (12) and (13) and the figure, it is found that the changing trends of recall rate and accuracy rate are similar. Therefore, in the following experiments, only one of the two needs to be selected as the index of the evaluation accuracy. This paper selects the recall rate of the recommended results as the index of the evaluation accuracy.

Use the K-means clustering algorithm to divide all users into 2, 3, 4, 5, and 6 clusters, and apply the recommendation algorithm in each cluster to recommend each item, and finally calculate the average recall rate and coverage rate. Table 1 is the recall rate of the recommended results obtained by the improved algorithm when the number of clusters is 2, 3, 4, 5, and 6. When the number of clusters is 1, the improved algorithm is an item-based collaborative filtering recommendation algorithm after an improved item similarity measurement method. [Table 2] shows the coverage of the recommended results in each clustering situation.

Table 1. The recall rate of recommended results in each clustering situation

	step=5	step=10	step=15	step=20	step=25	step=30	step=35	step=40	step=45
Item-CF					0.113607				
P=1	0.114562				0.114562				
P=2	0.115214	0.115991	0.115914	0.115914	0.115914	0.115914	0.115914	0.115914	0.115914
P=3	0.114872	0.115539	0.115264	0.11497	0.115213	0.11504	0.115313	0.115313	0.115362
P=4	0.112276	0.11376	0.113544	0.113991	0.114219	0.114026	0.113499	0.1138	0.113933
P=5	0.113462	0.113232	0.113434	0.11353	0.113348	0.113383	0.112076	0.113544	0.112922
P=6	0.11181	0.112852	0.11144	0.111783	0.111559	0.111797	0.111287	0.111363	0.112034

Table 2. Coverage of recommended results in each clustering situation

	step=5	step=10	step=15	step=20	step=25	step=30	step=35	step=40	step=45
Item-CF					0.307439				
P=1					0.296306				
P=2	0.263664	0.273785	0.274291	0.274291	0.274291	0.274291	0.274291	0.274291	0.274291
P=3	0.265688	0.292763	0.286184	0.289727	0.286184	0.284919	0.286184	0.286184	0.293522
P=4	0.302632	0.283907	0.285678	0.297571	0.28163	0.286943	0.281883	0.302379	0.307945
P=5	0.272267	0.276822	0.292763	0.292004	0.292257	0.283148	0.284919	0.290233	0.290992
P=6	0.291498	0.292763	0.291245	0.29504	0.293522	0.302632	0.281377	0.300101	0.294534

[Figure 3] shows the recall rate of recommended results. The abscissa indicates the number of iteration steps of the clustering algorithm, and the ordinate indicates the recall rate of the recommended results. When the number of iteration steps is about 10, the recall rate of the recommended results in various situations takes the maximum value. And when the user cluster class is 2, the number of iteration steps is about 10 steps, the algorithm has reached convergence, and the effect is also the best, and the recall rate has reached zero. 115 991, an improvement over the collaborative filtering recommendation algorithm 2. 1%. When all users are clustered into 2 or 3 clusters, the recommendation result of the top-N collaborative filtering recommendation algorithm based on the K-means clustering algorithm is higher than that of the top-N recommendation algorithm. When the number of clusters is 4, in most cases, the recall rate of the recommended results is also better than the original algorithm. When the number of clusters is 1, the item-based collaborative filtering recommendation algorithm after the item similarity measurement method is improved, the recommendation results obtained are better than the traditional collaborative filtering recommendation algorithm. The experimental results verify the original hypothesis of this article, that is, a user group that is more similar to the target user has a more valuable reference role than other users. Increasing the reference role of interest-related users to target users can improve the accuracy of recommendations for target users. As the number of clusters increases, the effect of the top-N collaborative filtering recommendation algorithm based on the K-means clustering algorithm becomes worse. As the number of clusters increases, the number of users in each cluster will inevitably decrease, that is, less user information can be referred to during recommendation, which will eventually lead to poor recommendation effects.

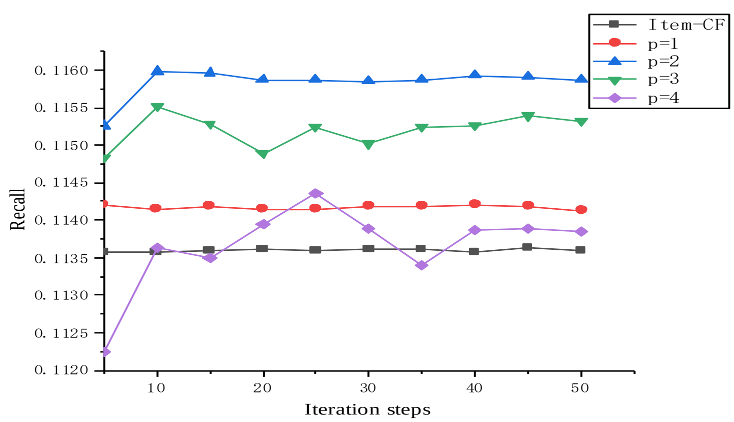

Figure 3. Recall rate in various situations

5. Conclusion

The traditional collaborative filtering recommendation algorithm refers to the purchase history information of all users when recommending a certain target user. However, when recommending for a target user, users who are more similar to it have a more valuable reference role than other users. Based on the item-based top-N collaborative filtering recommendation algorithm, this paper proposes a top-N collaborative filtering recommendation algorithm based on the K-means clustering algorithm. Using K-means to cluster according to the distance between the sample points, the similarity between users is regarded as the distance, the users are clustered into several clusters, and then the recommendation algorithm is applied in each cluster to perform recommendations. In addition, the traditional similarity measurement method only uses the user rating matrix to calculate the similarity between users or items, ignoring the difference in user ratings and the attributes of the items themselves. This article introduces item time difference and popular items in the similarity measurement formula. Factors such as weights and the weights of users' common ratings. The experimental results show that the recall rate of the algorithm proposed in this paper is higher than that of the traditional collaborative filtering recommendation algorithm2. 1%, proved the correctness of the proposed algorithm.

References

[1] M. K. Najafabadi, A. Mohame, and C. W. Onn, "An impact of time and item influencer in collaborative filtering recommendations using graph-based model," Information Processing and Management, vol.55, no.3, pp.526-540, (2019)

[2] H. Zarzour, Z. Al-Sharif, and M. Al-Ayyoub, "A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques," 2018 9th International Conference on Information and Communication Systems (ICICS), IEEE, (2018)

[3] L. I. Hongmei, W. Hao, and G. Chen, "Collaborative filtering recommendation algorithm based on exact Euclidean locality-sensitive hashing," Journal of Computer Applications, (2017)

[4] B. M. Sarwar, J. Konstan, and J. Riedl, "Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering," Proc of the 5th International Conference on Computer and Information Technology, (2002)

[5] A. Kelil, S. Wang, and R. Brzezinski, "CLUSS2: An alignment-independent algorithm for clustering protein families with multiple biological functions," International Journal of Computational Biology and Drug Design, vol.1, no.2, pp.122, (2008)

[6] Samara, Petros, Schomaker, "A multiple-label guided clustering algorithm for historical document dating and localization," IEEE Transactions on Image Processing, (2016)

[7] K. Mahdavi, "A clustering genetic algorithm for software modularization with a multiple hill-climbing approach," Brunel University, (2005)

[8] R. Steinberger, B. Pouliquen, and J. Hagman, "Cross-lingual document similarity calculation using the multilingual thesaurus EUROVOC," Springer Berlin Heidelberg, (2002)

[9] H. Ishii, S. Obata, and N. Niitsu, "Charge mobility calculation of organic semiconductors without the use of experimental single-crystal data," Scientific Reports, vol.10, no.1, (2020)

[10] G. Arora, A. Kumar, G. S. Devre, "Movie recommendation system based on users' similarity," International Journal of Computer Science and Mobile Computing, vol.3, no.4, pp.765-770, (2014)

[11] M. T. Puth, M. Neuhauser, G. D. Ruxton, "Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits," Animal Behavior, vol.102, pp.77-84, (2015)

[12] J. J. Castro-Schez, R. Miguel, and D. Vallejo, "A highly adaptive recommender system based on fuzzy logic for B2C E-commerce portals," Expert Systems with Applications, vol.38, no.3, pp.2441-2454, (2011)

[13] C. Anderson, "The long tail," S. L.: Wereldbibiotheek, (2013)

This page is empty by intention.