Research Article

Improved CF based prediction technique for recommendation systems

Ruchita Sharma[1*] and Manish Sharma[2]
M.Tech Scholar, Suresh Gyan Vihar University Jaipur, India[1]
Associate Professor, Suresh Gyan Vihar University Jaipur, India[2]

Abstract
Any eCommerce websites use recommendation system to recommend items to users. Collaborative filtering is a technique to recommend an item to the customers by understanding the past behavior of the same user and other similar users. The accuracy of the recommendation system is a major issue while recommending items to users. In this paper a research is done to propose a new system to predict ratings for items for the users by finding similarity between the users. Similarity between the users is found by analyzing the previous history of the users for rating items. A similarity matrix is created that store a similar weight between users. Similar users are selected if the similarity weight between the users is found greater than a similarity threshold. The proposed system is implemented on a data set and the quality of the proposed system is analyzed by comparing the value of mean absolute error (MAE). The experimental results are found better than some other existing techniques. The value of MAE is approximate 11% better and value of RMSE is 15% better as compared to existing algorithms.

Keywords
Recommendation systems, Collaborative filtering, Prediction system, MAE.

1.Introduction
The objective of the Recommendation System (RS) is to suggest items for the customers in whom he/she may be interested [1, 2]. Recommendation systems mainly examine the past history of the user for recommending items, in which he/she may be highly interested. Recommendation systems are very useful for eCommerce websites which are offering products online. When a new user is entered in the eCommerce website and shows interest to purchase an item then recommendation system suggests more items in which he/she is interested. The process of recommending some new items to the user improves the sale for e-commerce websites [3−8]. Recommendation system also worked as an information retrieval system for its users. It suggests items for the users in which he/she may be interested. It suggests the most appropriate items to its users on the basis of past behavior of its users [9, 10]. Recommendation systems are classified into three main categories on the basis of the method they are using for suggesting most appropriate items for users [1, 11].

The three main categories are (1) Content based filtering (2) Collaborative Filtering and (3) Hybrid Filtering.

Content based filtering mainly utilizes the past history of the user for the recommendation. It only utilizes the history of the same user and discards all the history of other users. It only checks the content of the same user. It recommends those items in which user shows interest in the past. Collaborative filtering (CF) is most commonly used technique by the recommendation systems [6, 7, 12, 13]. It uses the history of the other users to recommend items to a user. It calculates the similarity between users and recommend items, in which other users of same category shows interest in the past. CF is used by many of the e-commerce website like Amazon to recommend items to its users. For example, if a user shows interest in buying mobile of Samsung then Samsung mobile can be recommended to all users which have similar taste. CF is a very useful technique but it can only work when a sufficient amount of data is available to calculate similarity between the users.

---

*Author for correspondence

Many researchers are working in this area and many hybrid algorithms for, suggested in recent years [1]. These hybrid algorithms make use of both, content based filtering and collaborative filtering to suggest items to users.

In this paper a recommendation system using collaborative filtering is proposed. Next section discussed the literature survey of recent work in this area. Then section III is showing the proposed work and illustrated that how the proposed collaborative filtering-based recommendation system will work. Section VI shows the details of the results obtained by implementing the proposed system and analysis of these results. Then in the last conclusion of the work is presented.

The objectives of this paper are as follows:
1. To propose a novel approach for prediction of ratings of items.
2. To implement the proposed approach on a data set having one lacs ratings of items.
3. To compare the performance of the proposed approach by implementing it and comparing the results with other existing approaches.

2.Literature survey

This section gives a brief survey of the recent state of art in recommendation systems. Mohamed et al. [1] proposed recommendation system based on an adaptive genetic algorithm. The adaptive genetic algorithm was used to improve the prediction accuracy of recommendation system. Multi criteria recommendation system was used. The genetic algorithm was applied to a data set of users and movies. The adaptive genetic algorithm shows better results as compared to other techniques. Hassan and Hamada [2] analyse the performance of a recommendation system which was based on neural networks. The prediction accuracy of the recommendation system was checked on the basis of 7 different factors. The outcome of the analysis results that neural network-based recommendation system gives better results. Ze and Dengwen [3] proposed a recommendation algorithm on the basis of rating consistent. The algorithm uses common score values to relate users to each other and recommend the ratings on the basis of the similar users. Rodrigues et al. [4] proposed a recommendation system that deal with problem of user cold start and item cold start. If do not have history of the user, then this problem is known as user cold star problem. If the history of the item is not available, then this problem is known as item cold start. Rodrigues et al.

not only proposed a hybrid recommendation system on item-based and user-based collaborative filtering but also deal with user cold start and item cold start problems.

Ying and Cao [5] proposed a recommendation system that uses FCM based clustering to find similarity between the users. The method uses slope one algorithm for recommendation of the items. By using FCM based clustering algorithm the prediction performance of the slope one recommendation algorithm was improved. Gupta and Gadge [6] used item based collaborative filtering technique to recommend items. To find similarity between the users, the demographic based clustering algorithm was used. The hybrid system improves the prediction performance of the recommendation system. Wei et al. [7] proposed a new recommendation system which was based on collaborative filtering. Item based clustering was used to find similarity between the users. A global similarity metric was used to find similarity between the users. Shambour et al. [8] proposed a personalized recommendation system. It uses item-based CF technique which was based on multi criteria. Hassan and Hamada [9] proposed a recommendation system based on neural network. Multi criteria recommendation was applied that give prediction to the items-based ion the learning capability of the neural network. Adomavicius and Kwon [10] also proposed a recommendation system for multi criteria rating system. Konstan and Riedl [11] proposed a recommendation system that uses user experiences to predict ratings of the items. Zhu et al. [12] proposed a hybrid recommendation system that was based on case-based reasoning and a CF technique which was based on user data. Zarzour et al. [13] proposed a recommendation system. The system was based on clustering technique to make cluster of the users. The system performs dimensionality reduction that improves the quality of the prediction. Zhang et al. [14] presented an item-based filtering that used trusted enrich approach. The system uses user social similarities to suggest items to the user.

After going through this literature, it is observed that many techniques are already proposed by researchers to improve the prediction capability of the recommendation system. But the accuracy of these systems is not perfect and there is a need to work more on these systems to improve their performance. This paper introduces a new recommendation system that predicts the ratings of the users for the movies. The system work on a data set of users and movies.

The system first cluster users on the basis of the rating similarity between them and then uses the information about similar users to predict rating of items. Next section illustrates the proposed work.

3.Materials and methods

This paper proposed new recommendation systems that predict the ratings of the movies for users. The system work on a Movie Lens data which is available online [15]. The data set contains rating of 671 users, which rate 163949 movies on a scale from 0.5 to 5.0. The source of this data set is given at reference no [15]. The name of the repository is Movie Lens [15] which provides many data sets of the movies viewed by users. The systems first find the list of related users and then predict the rating for a given item for a given user by using related users. The two steps are explained here.

Finding related users: In this step a list of related users for a given user is formed. If two users give approximately same ratings to items, then these users have the same taste and will be a member of the list of related users of each other. For calculating related users, the ratings given by two users for items are analyzed in detail. If two users are rating a set of items, then a weight_of_relativeness between two users is calculated. For calculating weight_of_relativeness between two users, the sum of the absolute differences between the ratings of the items which are rated by both the users is calculated. If two users give equal ratings to all the items, then the value of that sum will zero and it shows maximum relativeness between two users. If weight_of_relativeness between two users is greater than a threshold, then these users will be the member of related users list of each other.

Prediction of ratings for a user-item pair: In this step, the rating for a user's item pair is calculated. We assume that the ratings for the given user item pair is not available in the data set. To predict rating for the given user, item pair, the average of non-zero ratings by all the related users of the given item is calculated. This average will be the predicted rating for the given

user item pair. Let us take a sample data of 10 users, which give ratings to five items on a scale of 1-10.

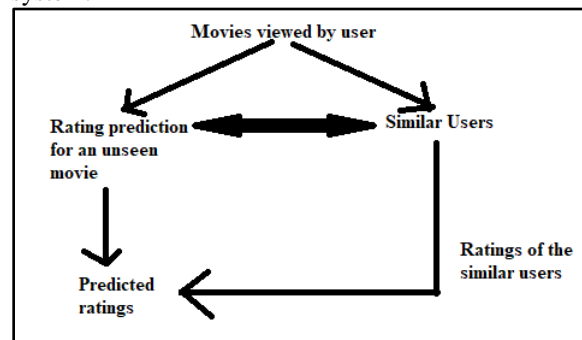*Figure 1* shows the block diagram of the proposed system.

Figure 1 Block diagram of the proposed system

The system takes the input from the data set about movies viewed by users. It predicts the ratings of an unseen movie for a user. It first calculates a list of related users for it. Then it finds the average ratings of the similar users for the given movie for which ratings are to be predicted. This average will be the predicted rating.

The Algorithm-1 shows the steps of the proposed work. It predicts the ratings of a movie MOV for the user U.
Algorithm-1
Input: ratings from data set, MOV, U
Output: Predicted Ratings for movie MOV by user U
Step 1: Read the ratings for different movies for different users from the data set.
Step 2: Find the related users for every user.
Step 3: Follow the steps from 4 to 6 to predict ratings of a movie MOV for a user U.
Step 4: Find a list RELATED_USR_LIST of related users for user U
Step 5: Calculate average AVG of ratings for movie MOV for every user in RELATED_USR_LIST.
Step 6: Return AVG as predicted ratings for movie MOV for user U.

Table 1 Ratings by 10 users for five items

S. No	Users	Item1	Item2	Item3	Item4	Item5
1	user1	7	6	8	7	10
2	user2	8	6	5	4	9
3	user3	4	8	4	1	9
4	user4	6	5	7	4	10
5	user5	4	7	8	4	3
6	user6	7	9	1	5	6

S. No	Users	Item1	Item2	Item3	Item4	Item5
7	user7	4	2	3	10	9
8	user8	5	6	5	3	3
9	user9	5	5	6	9	5
10	user10	5	5	6	5	5

Table 2 shows the difference between the sum of the ratings given by user-1 and all other 9 users i.e. from user2 to user9. Sum of absolute differences is calculated. Larger will the sum less will the similarity. The value of Max is the largest difference between the ratings of two users. In this case the maximum ratings for an item are on the scale 1-10 and the largest difference will be 10-1 = 9. The maximum value of sum that possible is (number of items rated × 9) =45. The similarity weight will be the 45-sum. The percentage of similarity determines the amount of similarity between the user1 and all other 9 users for this sample data.

Table 2 Showing the difference between the sum of the ratings given by user-1 and all other users

S. No	User	Item1	Item2	Item3	Item4	Item5	Sum	Similarity weight (Max-sum) = 45-Sum	similarity %age
1	user2	1	0	3	3	1	8	37	82.22
2	user3	3	2	4	6	1	16	29	64.44
3	user4	1	0	1	3	0	5	40	88.88
4	user5	3	1	0	3	7	14	31	68.88
5	user6	0	3	7	2	4	16	29	64.44
6	user7	3	4	5	3	1	16	29	64.44
7	user8	2	0	3	4	7	16	29	64.44
8	user9	2	1	2	2	5	12	33	73.33
9	user10	2	1	2	2	5	12	33	73.33
10	user2	1	0	3	3	1	8	37	82.22

Let similarity threshold is 75%. So, if the value of the similarity percentage between user1 and other users is greater than 75 then those users will be the member of similarity list of user1. How much should be the similarity threshold will be the future scope of this work. So user1 similarity list will contain user2 (with similarity %age 82.22%) and user4 (with similarity 5age 88.88%).
User 1 similarity list = {user2, user4}
Let us predict ratings for user1 for the item item2.
Given user= user1
Given item = item2
Sum of Ratings for item2 by related users = 6   (by user2) + 5 (by user4) = 11
Average ratings = (11/2) = 5.5
So predicted rating for item2 by user1 is 5.5.
The proposed method is implemented to predict the ratings of user item pairs for a data set available online at [15]. The next section discussed the results of the proposed work after implementation.

4.Results
The proposed work discussed in the previous section is implemented on a data set of 20 users. In the data set user predict ratings for movies on a scale from 0 to 5. We select a sample of 50 user item pairs and predict their ratings on the basis of similarity between users. Out of 50 selected user item pairs it may be the case that the similar user does not predict the same item for which the rating is to be proposed. In the final result the ratings of those similar users are entertained, which rate the same item for which we are predicting the rating. The snapshot in the figure is showing the actual ratings and predicted ratings of the items.

5.Discussion
The results obtained by experiments are evaluated on two metrics MAE and RMSE which are as follows.
Mean absolute error (MAE): MAE was used to measure our prediction accuracy; it measures how close our predicted ratings are to the actual outcome.

Root mean square error (RMSE): RMSE was also used to measure the prediction accuracy, it uses the squared deviation and emphasizes on large errors.

The value of the MAE is 0.58 and value of RMSE is 0.68 using the proposed algorithm. *Table 1* compares the performance of proposed work with some other

recent works found in literature. The best of the value of MAE is 0.65 and best values of RMSE is 0.80 found by [15]. The value of MAE is 10.7% better and value of EMSE is 15% better as compared to the value found by [15]. *Table 3* shows the comparison of results of the proposed work with recent works

Table 3 Comparison of results of proposed work with recent works

Metric	Adaptive GA [1]	NN-Based Recommender systems [2]	FCM and slope one algorithm [5]	Proposed work	%age of improvement as compare to [15]
MAE	1.59	1.52	0.65	0.58	10.70%
RMSE	2.12	2.53	0.8	0.68	15%

The paper fulfills its all the objectives specified. It proposed a new approach for prediction of ratings, implemented it and compare the results with other algorithms. The results obtained are found better.

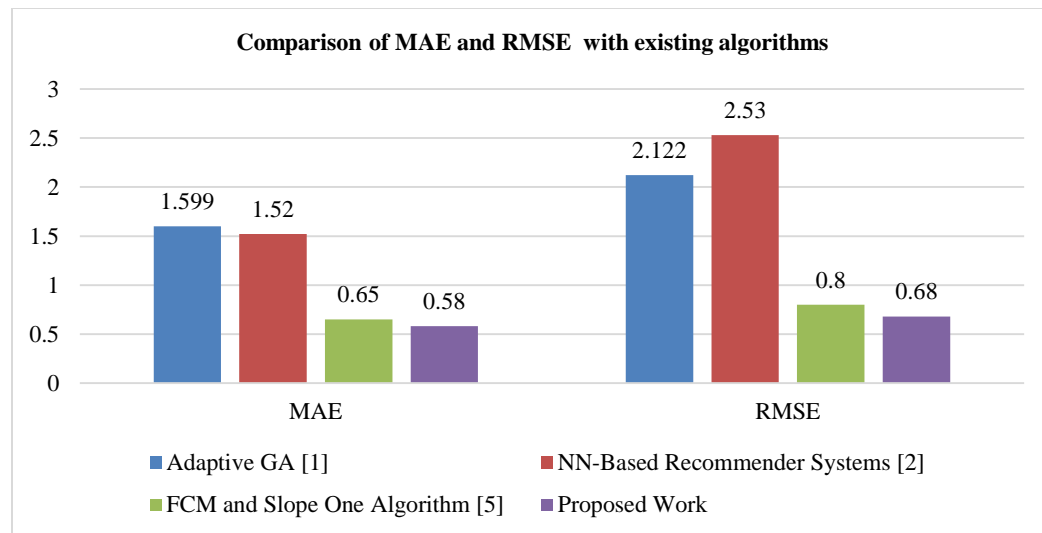*Figure 2* shows a graph which compares the results of existing algorithms and proposed algorithm.

Figure 2 Result comparison of MAE and RMSE

6. Conclusions and future scope

This paper proposed a novel approach for predicting ratings of items for recommendation systems. These systems help to recommend items to users on e-commerce web sites. A novel approach to predict ratings of movies for users. The implementation results suggest that the system producing good results. The following has been concluded:

(1) User based recommendation systems perform better in predicting ratings of items
(2) Finding similarity between users according to their interest is very useful in prediction systems.

The future scopes of this work are as follows:

1. In future this approach can be tested on other data sets to validate its performance
2. Improvements in finding similarity between users can be made by using clustering algorithms such as K-Means etc.

References
[1] Mohamed H, Abdulsalam L, Mohammed H. Adaptive genetic algorithm for improving prediction accuracy of a multi-criteria recommender system. In international symposium on embedded multicore/many-core systems-on-chip 2018 (pp. 79-86). IEEE.
[2] Hassan M, Hamada M. Performance analysis of neural networks-based multi-criteria recommender systems. In international conferences on information technology, information systems and electrical engineering 2017 (pp. 490-4). IEEE.
[3] Ze W, Dengwen Z. Optimization collaborative filtering recommendation algorithm based on ratings

consistent. In international conference on software engineering and service science 2016 (pp. 1055-8). IEEE.

[4] Rodrigues CM, Rathi S, Patil G. An efficient system using item & user-based CF techniques to improve recommendation. In international conference on next generation computing technologies 2016 (pp. 569-74). IEEE.

[5] Ying Y, Cao Y. Collaborative filtering recommendation combining FCM and slope one algorithm. In international conference on informative and cybernetics for computational social systems 2015 (pp. 110-5). IEEE.

[6] Gupta J, Gadge J. Performance analysis of recommendation system based on collaborative filtering and demographics. In international conference on communication, information & computing technology 2015 (pp. 1-6). IEEE.

[7] Wei S, Ye N, Zhang S, Huang X, Zhu J. Collaborative filtering recommendation algorithm based on item clustering and global similarity. In international conference on business intelligence and financial engineering 2012 (pp. 69-72). IEEE.

[8] Shambour Q, Hourani M, Fraihat S. An item-based multi-criteria collaborative filtering algorithm for personalized recommender systems. International Journal of Advanced Computer Science and Applications. 2016; 7(8):274-9.

[9] Hassan M, Hamada M. A neural networks approach for improving the accuracy of multi-criteria recommender systems. Applied Sciences. 2017; 7(9):1-18.

[10] Adomavicius G, Kwon Y. New recommendation techniques for multicriteria rating systems. IEEE Intelligent Systems. 2007; 22(3):48-55.

[11] Konstan JA, Riedl J. Recommender systems: from algorithms to user experience. User Modeling and User-Adapted Interaction. 2012; 22(1-2):101-23.

[12] Zhu X, Ye H, Gong S. A personalized recommendation system combining case-based reasoning and user-based collaborative filtering. In Chinese control and decision conference 2009 (pp. 4026-8). IEEE.

[13] Zarzour H, Al-Sharif Z, Al-Ayyoub M, Jararweh Y. A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques. In international conference on information and communication systems 2018 (pp. 102-6). IEEE.

[14] Zhang H, Ganchev I, Nikolov NS, O'Droma M. A trust-enriched approach for item-based collaborative filtering recommendations. In international conference on intelligent computer communication and processing (ICCP) 2016 (pp. 65-8). IEEE.

[15] https://grouplens.org/datasets/movielens/. Accessed 12 March 2019.

Ruchita Sharma pursuing M.Tech from Suresh Gyan Vihar University, Jaipur. She is working on optimizing the performance of recommendation systems using collaborating filtering. Her area of research includes Data Mining and Recommendation Systems.

Email: ruchita.sharma08@gmail.com

Manish Sharma is working as Associate Professor at Suresh Gyan Vihar University, Jaipur. He is an active researcher in Computer Science and Engineering and published many papers in International Journals and Conferences. His current area of research is Recommendation Systems using Collaborative Filtering.