

Recommender system using item based collaborative filtering (CF) and K-means

Mamata Garanayak^{a,*}, Sachi Nandan Mohanty^b, Alok Kumar Jagadev^b and Sipra Sahoo^c

^aDepartment of Computer Science, CUTM, State Private University, Bhubaneswar 752054, India

^bSchool of Computer Engineering, KIIT, Deemed to be University, Bhubaneswar 751031, India

^cDepartment of Computer Science, SOA, Deemed to be University, Bhubaneswar 751030, India

Abstract. The heightening in the available information in the form of digital data and the number of users on the Internet have engendered a challenge of overburden of data which obstructs access to interested item on the Internet timely. There are many information retrieval systems which try to solve the problem of information overloading but in their cases prioritization and personalization of information were absent. The main aim is to develop a recommender system using item based collaborative filtering technique and K-means. The most popular algorithm in the recommender system's field is the collaborative filtering technique. Recommender systems are the filtering systems for information that concerned with the problem of information overburden by filtering essential information fragment out of enormous dynamically promoted information according to person's attentiveness, taste and distinguished behavior about them. We are considering m users, n items (in numbers) and presenting a model to fabricate a recommendation for the mobile user by a new approach.

Keywords: Recommender system, item based collaborative filtering, K-means clustering, data mining

1. Introduction

Normally almost a huge amount of data is available in the Internet and also the number of users in the Internet increases rapidly. The user on the Internet has no time to search everything on the Internet due to their busy schedule. In this era of competition, information causes overloading which in turn are time consuming. So it is quite necessary to recommend the items to the users based on their interest and preferences. A recommender system plays a vital role in such field.

Any system that manufacture individualized recommendations as output or has the consequence of steering the user in a personalized way to interesting or useful items in a large space of possible options are called as the recommender system. These systems are the brand new proficiency of promotion of movie, music, home products, electronic items and all the things that we use in our day to day life. The manufacturer

and suppliers had difficulty in offering products that fulfill the customer criteria of buying the products.

Recommender systems play an important role for both the Internet users and service providers. It decreases the proceeding costs of discovery and adopts item in a territory of online shopping.

2. Literature review

2.1. Recommendation system

Recommender systems are capable to recognize whether a specific shopper would like a specific product or not, based on the user's profile [8].

To pick a book from a set of choices is easier but when the set of choices is as large as a library then a recommendation system comes to the picture.

A recommender system is a system to which we give a set of inputs, apply a suitable algorithm and provide the output as a recommendation item as per the user choices and preferences. Here the input data are the set of items across which recommendation might be

*Corresponding author: Mamata Garanayak, Department of Computer Science, CUTM, State Private University, Bhubaneswar, Odisha, 752054, India. E-mail: mamatagaranayak@gmail.com.



Fig. 1. Example of recommender system.

constructed (I), set of users whose proclivity are well known (U), users for whom recommendations required to be created (u), and items for which we would like to forecast u's proclivity and the output is the u's predicted preference for [11].

A recommender system has 4 parts.

1. Database where the inputs data are available
2. An interface like computer
3. Algorithm
4. Recommendation component as an output

2.2. Data mining

Mining of data is the mining of knowledge from data i.e. extricating serviceable information from the crude data. The techniques by which the mining of data occurs include clustering of the sets of data points, categorization of data, prediction of data, decision tree, link analysis, outlier detection, association rules, sequence analysis, time series analysis and text mining, and also some up to the minute techniques such as sentiment analysis and analysis of social networks [6,16].

The techniques of data mining are the outcome of a prolonged research and product expansions or evolutions [14]. The expansion started when a enormous amount of business aspects was first cached on systems, sustained with refinement in access of the data, and more recently provoke technologies that allow persons to steer through their sets of data in an environment such as factual hour. Data mining captures the evolutionary action beyond focus back the data process, access and navigation to prospective and proactive delivery of data. Data mining appeal is organized for the circle such as business as it is bear by three technologies:

- Collection of enormous data
- Strong multiprocessor systems
- Mining algorithms for data sets

The mining of data occurs in 3 steps: (1) Initial exploration. (2) Erection of model and validation. (3) Deployment.

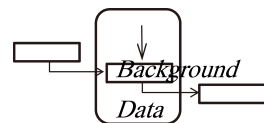


Fig. 2. Recommender system.

2.2.1. Stage 1: Exploration

This stage normally go ahead with data construction which involves data polish, data conversion, and assortment of the records and - in case of data sets with huge numbers of variables or fields- accomplishing few key feature selection operations to bring the numbers of variables to an achievable range [14]. Then, depending on the features of the problem, the initial phase of the process of the data mining may involve anywhere between a simple choice of simple predictors for a regression model, to elaborately describe the analysis using a broad variation of statistical and graphical techniques (such as Exploratory Analysis of Data) in order to recognize the most alike variables and regulate the complexity and/or the common characteristics of models that can be griped into the succeeding phase.

2.2.2. Stage 2: Erection of model and validation

Model erection and validation stage contemplate the different models used in mining of data and adopt the leading one ground on their performance (i.e., it describes the query's variability and acquire the steady outcomes over the sample data sets). This may noise as a simplest working, but really, it occasionally necessitates a very detailed and prolonged procedure. There are different expertise to reach these type of objectives – many of them are relies on normally called “competitive evaluation models”, that is, applying different models to the same data sets and then comparing their performance to pick the foremost among them. These methods are also called as the basics of predictive data mining in order to lessen the variance includes: Bagging (Voting, Averaging), Meta-Learning, Stack Generalizations (Stacking), and Boosting [10,15]. Validation is the process of how better the mining models performs against the actual data.

2.2.3. Stage 3: Deployment

Deployment is the final and last stage of mining of data which includes:

- Selection of one model as best in the model building stage.
- Apply the best model to the up to date data set in sequence to produce expected outcome as a result.

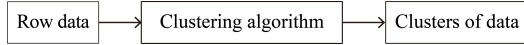


Fig. 3. Clustering.

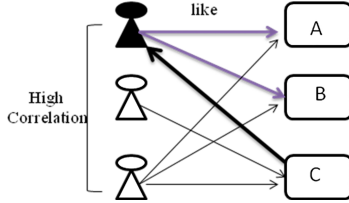


Fig. 4. User based collaborative filtering.

2.3. Clustering

Clustering or cluster inspection is the job of assemble the deck of data points in such a method that the data points in the identical category (termed as a cluster) are more identical (in few cases) to each other than to those which are in different assemble. It is the vital job of mining of data, and a customary method for statistical inspection of data, worn in numerous fields including study of the machine, pattern identification, image inspection, compression of data, and retrieval of information and computer graphics.

In simple words, the aim of clustering is to divide assembles with alike attributes and allocate them into clusters. Clustering is split into two sub categories:

- **Hard Clustering:** In hard clustering, each data either belongs to one cluster entirely or not belongs to that cluster.
- **Soft Clustering:** In soft clustering, a possibility or chance of the data point to be in those clusters is allocated instead of placing each data point into a distinct cluster.

A Clustering Algorithm tries to analyze the groups of data on the basis of similarity. It found the centroid of the group of data points. To carry out clustering, the algorithm evaluates the distance between each point from the centroid of the cluster. The principal focus of clustering is to dictate the inherent grouping in sets of data that are unlabelled.

2.4. K-means

K-mean clustering is an unsupervised learning algorithm, which is well liked for cluster inspection in data mining. It focuses on separation of N numbers of data into K number of clusters where each data belongs to the cluster with proximate mean distance.

Algorithm:

-
- Step 1: Initialize the cluster center.
 - Step 2: Compute the distance between the points using distance function $P(a, b) = |x_2 - x_1| + |y_2 - y_1|$.
 - Step 3: Allocate the data to the cluster whose interval from the cluster nucleus is minimum of the entire nucleus of the cluster [17].
 - Step 4: Upgrade the nucleus of the cluster.
 - Step 5: Re-compute the distance from each data points and newly acquire cluster nucleus.
 - Step 6: If there is no new data stop or else go to the step 3 [5].
-

2.5. Collaborative filtering

Collaborative filtering perceives a subset of persons who have similar flavour and preferences to the target person and apply this subset for recommendations [3]. It is commonly categorized into 2 types.

- a. Model based collaborative filtering
- b. Memory based collaborative filtering

Model based collaborative filtering techniques inspects the user-item matrix to identify relation among the items; they apply these connections to differentiate the lists of recommendations [9,12].

An example of these techniques includes clustering, regression, decision tree, Link analysis, etc.

Memory based collaborative filtering is categorized into 2 types.

- a. Collaborative filtering based on users
- b. Collaborative filtering based on items

2.5.1. Collaborative filtering based on users

In this technique recommendations are stated to the users based on the consideration of items by other users from the similar group, with whom he/she shares customary preferences [1,2].

User correlation:

$$UserSimilarity(u, m) = \frac{\sum_{i \in CR_{u,m}} (r_{ui} - r_u^-)(r_{mi} - r_m^-)}{\sqrt{\sum_{i \in CR_{u,m}} (r_{ui} - r_u^-)^2} \sqrt{\sum_{i \in CR_{u,m}} (r_{mi} - r_m^-)^2}} \quad (1)$$

Where

$UserSimilarity(u, m)$ = similarity among two users u and m .

r_{ui} = rating of item i given by the user u .

r_u^- = mean rating of the user u .

r_{mi} = rating of item i given by the user m .

r_m^- = mean of the user m .

Prediction function:

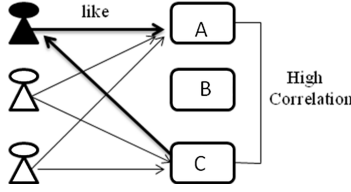


Fig. 5. Item based collaborative filtering.

$$prediction(u, i) = r_u^- + \frac{\sum_{m \in n_{neighbor}(u)} UserSimilarity(u, m) \cdot (r_{mi} - r_m^-)}{\sum_{m \in n_{neighbors}(u)} UserSimilarity(u, m)} \quad (2)$$

2.5.2. Item based collaborative filtering

In this category the taste of person remains fixed or changes quite little. Alike items erect neighborhoods based on persons [7]. After that the system produces recommendations with items that a user would prefer in the neighborhood [2,4,13].

Item similarity:

$$itemSimilarity(i, j) = \frac{\sum_{u \in ucRB_{i,j}} (r_{ui} - r_{u1}^-)(r_{uj} - r_{u2}^-)}{\sqrt{\sum_{u \in ucRB_{i,j}} (r_{ui} - r_{u1}^-)^2} \sqrt{\sum_{u \in ucRB_{i,j}} (r_{uj} - r_{u2}^-)^2}} \quad (3)$$

Where

$itemSim(i, j)$ = similarity among two items i and j .

r_{ui} = rating of item i given by the user u .

r_{uj} = rating of item j given by the user u .

r_{u1}^- = mean rating of first item.

r_{u2}^- = mean rating of second item.

Prediction function:

$$prediction(u, i) = \frac{\sum_{j \in ratedItems(u)} itemSimilarity(i, j) \cdot r_{uj}}{\sum_{j \in ratedItems(u)} itemSimilarity(i, j)} \quad (4)$$

3. Methodology

As the methodology of recommendation system, K-means and item based collaborative filtering techniques steps are as follows:

3.1. Needs of the system

Studied the algorithm used for k-means and item based collaborative filtering techniques. Also study the need for the data which can be used in Table 1.

Table 1
The movies liking form rating scale

Liking	Rating scale
Do not like	★
Not like	★★
Normal	★★★
Like	★★★★
Very Like	★★★★★

Table 2
Synthetic data

User	Age	Rating
User 1	15	5
User 2	17	2
User 3	20	3
User 4	22	5
User 5	25	4
User 6	15	4
User 7	30	2
User 8	55	3
User 9	50	5
User 10	32	4
User 11	10	3.5
User 12	15	4
User 13	40	5
User 14	45	2.5
User 15	32	3
User 16	25	4.5
User 17	20	3
User 18	9	5
User 19	13	2.5
User 20	29	4
User 21	47	5
User 22	60	3
User 23	72	4.5
User 24	65	3.5
User 25	61	4.5
User 26	58	2.5
User 27	55	3
User 28	58	3
User 29	50	4
User 30	28	3.5
User 31	25	4
User 32	37	4.5
User 33	35	5
User 34	42	2.5
User 35	40	4
User 36	53	4
User 37	52	4.5
User 38	72	4
User 39	70	5
User 40	65	3.5
User 41	62	3.5
User 42	51	4
User 43	81	3.5
User 44	78	4
User 45	63	3.5
User 46	79	5
User 47	75	5
User 48	71	4
User 49	63	3.5
User 50	81	3.5
User 51	85	4

Table 3
Cluster formation

Users	Age	Rating	(15, 4) Cluster 1	(40, 5) Cluster 2	(65, 3.5) Cluster 3	(28, 3.5) Cluster 4	(51, 4) Cluster 5	(83, 3.5) Cluster 6	(71, 4) Cluster 7	Cluster
User 1	15	5	1	25	51.5	14.5	37	69.5	57	1
User 2	17	2	4	26	49.5	12.5	36	67.5	56	1
User 3	20	3	6	22	45.5	8.5	32	63.5	51	1
User 4	22	5	8	18	43.5	7.5	30	62.5	50	4
User 5	25	4	10	16	30.5	3.5	26	48.5	46	4
User 6	15	4	0	26	50.5	13.5	36	68.5	56	1
User 7	30	2	17	13	36.5	3.5	23	54.5	43	4
User 8	55	3	41	17	10.5	27.5	5	28.5	17	5
User 9	50	5	36	10	16.5	13.5	2	34.5	22	5
User 10	32	4	17	9	33.5	4.5	19	51.5	39	4
User 11	10	3.5	5.5	31.5	55	18	41.5	73	61.5	1
User 12	15	4	0	26	50.5	13.5	36	68.5	56	1
User 13	40	5	26	0	26.5	13.5	11.5	44.5	32	2
User 14	45	2.5	31.5	7	21	18	7.5	39	30.5	2
User 15	32	3	18	10	33.5	4.5	20	51.5	40	4
User 16	25	4.5	10.5	15.5	41	4	19.5	59	46.5	4
User 17	20	3	6	21	45.5	8.5	32	63.5	56	1
User 18	9	5	7	31	57.5	20.5	43	75.5	63	1
User 19	13	2.5	3.5	29.5	53	16	39.5	71	59.5	1
User 20	29	4	14	12	36.5	1.5	22	54.5	42	4
User 21	47	5	33	7	19.5	20.5	5	37.5	25	5
User 22	60	3	46	22	5.5	32.5	10	23.5	12	3
User 23	72	4.5	57.5	32.5	8	45	21.5	12	1.5	7
User 24	65	3.5	50.5	26.5	0	27	14.5	18	6.5	3
User 25	61	4.5	46.5	21.5	3	34	10.5	23	10.5	3
User 26	58	2.5	44.5	20.5	8	31	8.5	26	14.5	3
User 27	55	3	41	17	10.5	27.5	5	28.5	17	5
User 28	58	3	44	20	7.5	30.5	8	25.5	14	3
User 29	50	4	35	11	15.5	22.5	1	35.5	21	5
User 30	28	3.5	13.5	13.5	37	0	23.5	45	43.5	4
User 31	25	4	10	16	40.5	3.5	26	48.5	46	4
User 32	37	4.5	12.5	3.5	29	10	14.4	37	34.5	2
User 33	35	5	21	5	31.5	6.5	17	32.5	20	2
User 34	42	2.5	28.5	4.5	24	15	10.5	42	30.5	2
User 35	40	4	25	1	25.5	12.5	11	43.5	12.5	2
User 36	53	4	38	14	12.5	25.5	2	30.5	18	5
User 37	52	4.5	37.5	12.5	14	25	1.5	32	19.5	5
User 38	72	4	57	33	7.5	44.5	21	11.5	1	7
User 39	70	5	46	30	6.5	43.5	20	14.5	2	7
User 40	65	4.5	50.5	25.5	0	38	14.5	19	7.5	3
User 41	62	3.5	47.5	23.5	3	34	11.5	21	12.5	3
User 42	51	4	36	12	14.5	23.5	0	32.5	20	5
User 43	81	4.5	56.5	41.5	17	54	30.5	3	11	6
User 44	78	4	63	39	13.5	50.5	27	5.5	7	6
User 45	63	3.5	48.5	24.5	0	35	12.5	20	8.5	3
User 46	79	5	65	39	15.5	52.5	29	17.5	9	7
User 47	75	5	61	35	11.5	48.5	25	9.5	5	7
User 48	71	4	56	32	6.5	43.5	20	12.5	0	7
User 49	63	3.5	48.5	24.5	0	35	12.5	20	8.5	3
User 50	81	3.5	66.5	42.5	16	55	31.5	2	12.5	6
User 51	85	4	70	46	20.5	57.5	34	2.5	14	6

3.2. Data

This part shows the basic data used to develop the system for the group of users with K-means and set the data used to create a database of the system. For this here the synthetic data of 51 users has been considered.

3.3. Processing model for analysis on item recommendation (Fig. 6)

3.4. Proposed algorithm

Stage 1: Clustering

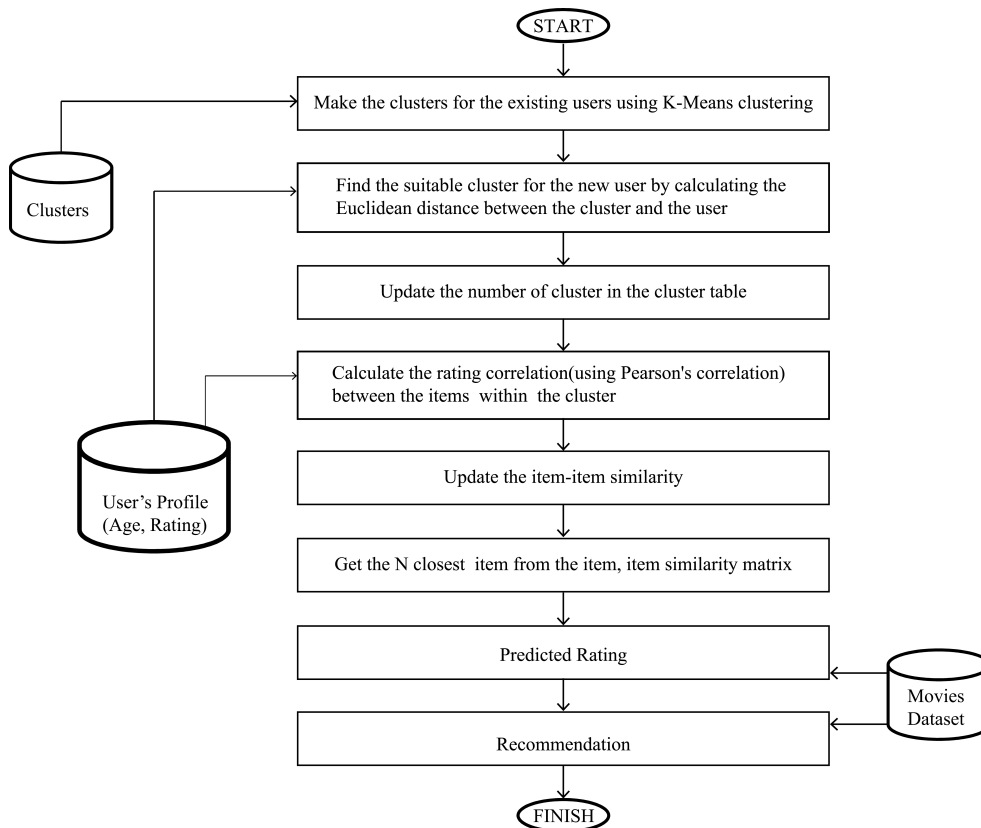


Fig. 6. Procedure for item recommendation.

Step 1.1 Randomly choose the N number of cluster centroid.

Step 1.2 Compute the distance using distance function $P(a, b) = |x_2 - x_1| + |y_2 - y_1|$.

Step 1.3 Allocate the user to the cluster whose distance from the centroid is minimum of the entire centroid.

Stage 2: Allocating a new user to the existing cluster

Step 2.1 Calculate the distance of new user from each centroid using Euclidean Distance.

Step 2.2 The user will enter to that cluster whose Euclidean distance from the user to the centroid is minimum.

Stage 3: Item based collaborative filtering

Step 3.1 Compute ItemSim (item i , item j) using Pearson's correlation for all the items if both item i and j are rated by the user.

Step 3.2 If there is a positive correlation, then that is taken into consideration.

Step 3.3 Calculate the prediction function

Table 4
Nature of the group is to divide with K-means

Clusters	Member of group
1	9
2	6
3	8
4	9
5	8
6	5
7	6

From the Table 3 it is found that, 7 clusters are formed and the nature of the group is to divide with k-means is as in Table 4.

3.5. Processing K-means and item based collaborative filtering

The system cluster with K-Means algorithm by calculating the distance of all points of data from the center of 7 groups by using Euclidean Distance and the information will be stored in database.

Table 5 shows the rating of the movies given by the user and the Table 6 shows the data of centroid for 4 movies.

Table 5
User gives movie rating

User ID	Movie ID	Rating
User 52	2858	4
User 52	2959	5
User 52	3243	3
User 52	3510	4

Table 6
Data of centroid for 4 movies

Movie ID	K_1	K_2	K_3	K_4	K_5	K_6	K_7
2858	4	5	2	3	5	2	3
2959	3	3	2	3	4	2	3
3243	2	3	4	4	3	4	2
3510	3	2	5	5	3	4	2

The distances of groups of users with $k_1, k_2, k_3, k_4, k_5, k_6$ and k_7 are:

$$D1 = \sqrt{(4-4)^2 + (3-5)^2 + (2-3)^2 + (3-4)^2} = 2.449$$

$$D2 = \sqrt{(5-4)^2 + (3-5)^2 + (3-3)^2 + (2-4)^2} = 3$$

$$D3 = \sqrt{(2-4)^2 + (2-5)^2 + (4-3)^2 + (4-5)^2} = 3.872$$

$$D4 = \sqrt{(3-4)^2 + (3-5)^2 + (4-3)^2 + (5-4)^2} = 2.645$$

$$D5 = \sqrt{(5-4)^2 + (4-5)^2 + (3-3)^2 + (3-4)^2} = 1.732$$

$$D6 = \sqrt{(2-4)^2 + (2-5)^2 + (4-3)^2 + (4-4)^2} = 3.741$$

$$D7 = \sqrt{(3-4)^2 + (3-5)^2 + (2-3)^2 + (2-4)^2} = 3.612$$

From the calculation it is found that, the users are separated from the least common group to the greater, so the system will provide the user 52 in fifth group.

After this, the system will search for an item similarity based on the item based collaborative filtering and will create a matrix of data between users and the movie rating given by the users as in Table 7.

In Table 7, the user 4 likes (gives rating of 5) the item 4 (i.e. Life of a Pie). Now the item based collaborative filtering technique is applied to find which item is similar to the item 4, so that, that item is recommended to the user 4.

Similarity between item 4 and item 1:

$$\text{Now } r_{u1}^- = (5 + 4)/2 = 3.5$$

$$r_{u2}^- = (3 + 2)/2 = 2.5$$

ItemSim (item 4, item 1) =

$$\frac{(4-3.5)(3-2.5) + (5-3.5)(2-2.5)}{\sqrt{(4-3.5)^2 + (3-2.5)^2} \sqrt{(5-3.5)^2 + (2-2.5)^2}} = 0$$

$$\text{Again } r_{u1}^- = (5 + 3)/2 = 4$$

$$r_{u2}^- = (2 + 5)/2 = 3.5$$

$$\frac{(5-4)(2-3.5) + (3-4)(5-3.5)}{\sqrt{(5-4)^2 + (2-3.5)^2} \sqrt{(3-4)^2 + (5-3.5)^2}} = 0$$

Similarity between item 4 and item 2:

$$\text{Now } r_{u1}^- = (5 + 3)/2 = 4$$

$$r_{u2}^- = (4.5 + 3.5)/2 = 4$$

ItemSim (item 4, item 2) =

$$\frac{(5-4)(4.5-4) + (3-4)(3.5-4)}{\sqrt{(5-4)^2 + (4.5-4)^2} \sqrt{(3-4)^2 + (3.5-4)^2}} = 0.8$$

The similarity between item 4 and item 3, item 4 and item 5, item 4 and item 7, item 4 and item 9 will not be calculated as user 4 has not rated that movie.

Similarity between item 4 and item 6:

$$\text{Now } r_{u1}^- = (4 + 5)/2 = 4.5$$

$$r_{u2}^- = (4.5 + 1.5)/2 = 3$$

ItemSim (item 4, item 6) =

$$\frac{(4-4.5)(4.5-3) + (5-4.5)(1.5-3)}{\sqrt{(4-4)^2 + (4.5-3)^2} \sqrt{(5-4.5)^2 + (1.5-3)^2}} = -0.6$$

Likewise the similarity among item 4 and item 8 can be calculated as 0.

The above calculation is similar to person correlation. Thus we found:

The similarity among item 4 and item 1 = 0.

The similarity among item 4 and item 2 = 0.8.

The similarity among item 4 and item 6 = -0.6.

The similarity among item 4 and item 8 = 0.

3.6. Prediction function

Now we calculate the prediction function. Here the size of the item is 3.

$$K = \{\text{item 1, item 2, item 8}\}$$

$$P_{u_4 I_4} = \frac{(0*2) + (0.8*4.5) + (0*2)}{|0+0.8+0|} = 4.5$$

That is, the item which is rated 4.5 by user 4 is similar to the item 4 i.e. item 2. So the item 2 is recommended to the user 4 as user 4 likes item 4.

4. Performance

The main aim of clustering is to know the number of people in the groups and the centroid of the group.

Table 7
Table of rating given by the users

	Jumanji	Money Train	Bahubali	Life of a pie	Casino	Firangi	Golmal	Sulu	Itihas
User 1	2	3.5	3	–	3	–	3	2.5	4
User 2	–	3	1.5	–	–	4	3	3.5	2
User 3	3	–	3	4	2.5	4.5	5	2	–
User 4	2	4.5	–	5	–	1.5	–	2	–
User 5	5	3.5	2	3	2	–	–	2	4
User 6	4	2	3	–	1.5	3.5	3	3	3

Table 8
Matrix of item 4 and item 1

Item 4	Item 1
–	2
–	–
4	3
5	2
3	5
–	4

Table 9
Matrix of item 4 and item 2

Item 4	Item 2
–	3.5
–	3
4	–
5	4.5
3	3.5
–	2

Table 10
Matrix of item 4 and item 6

Item 4	Item 6
–	–
–	4
4	4.5
5	1.5
3	–
–	3.5

Then bring the centroid to a cluster group for new user to the group by k-means algorithm. In this paper, we use R Studio software. It is used to cluster a group of users, the data downloaded from the website movie lens and here we consider 51 user's data.

5. Conclusion and future work

Out of all the recommendation system technique, collaborative filtering technique is the most popular one. In this paper the data is clustered using K-means clustering and after that item based collaborative filtering technique is used to recommend the most similar item to the particular user. In future instead of k-means clustering, fuzzy c-mean clustering can be applied and either collaborative filtering based on user or item col-

```

Console Terminal x
~/RStudio
> table(book1$Rating, ans$cluster)
 1 2 3 4 5 6 7
2 0 0 0 0 1 1
2.5 0 0 0 1 2 0 1
3 2 0 0 2 0 1 2
3.5 0 2 0 5 0 1 1
4 3 2 0 1 4 2
4.5 1 0 1 1 1 1 0
5 2 1 2 0 2 1 2
> plot(Age ~ Rating, col= ans$cluster, main = "Rating", pch=20)
> plot(Age ~ Rating, col= ans$cluster, main = "Rating", pch=15)
> plot(Age ~ Rating, col= ans$cluster, main = "Rating", pch=18)
> plot(Age ~ Rating, col= ans$cluster, main = "Rating", pch=20)
> plot(Age ~ Rating, col= ans$cluster, main = "Rating", pch=21)
> plot(Age ~ Rating, col= ans$cluster, main = "Rating", pch=16)

```

Fig. 7. Formation of cluster.

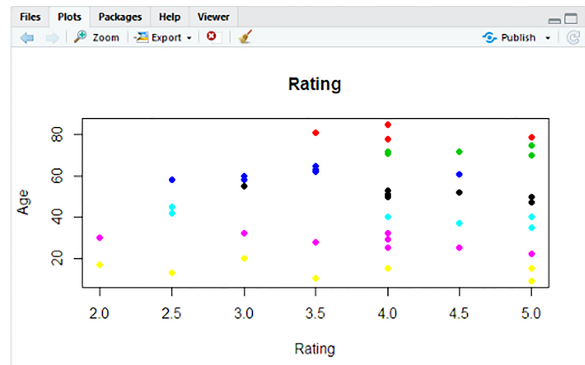


Fig. 8. Number of groups and members of the group.

laborative filtering based on item can be applied to recommend the best item to the user.

References

- [1] P. Phorasim and L. Yu, Movies recommendation system using collaborative filtering and k-means, *International Journal of Advanced Computer Research* 7(29) (2017), ISSN (Print): 2249–7277 ISSN (Online): 2277–7970. <http://dx.doi.org/10.19101/IJACR.2017.729004>.
- [2] S. Gong, A collaborative filtering recommendation algorithm based on user clustering and item clustering, *Journal of Software* 5(7) (2010), 745–752.
- [3] G. Adomavicius and A. Tuzhilin, Towards the next generation of recommender system: A survey of the state-of-the-art and possible extensions, *IEEE Transactions on Knowledge and Data Engineering* 17(6) (2005), 734–749.
- [4] C. Pan and W. Li, Research paper recommendation with topic analysis, in: *Computer Design and Applications IEEE* 2010, pp. V4–264.

- [5] A. Felfernig, G. Friedrich and L. Schmidt-Thieme, Guest editors' introduction: Recommender systems, *IEEE Intelligent Systems* **22**(3) (2007), 18–21.
- [6] J. Han and M. Kamber, *Data mining: Concepts and Techniques*, Elsevier; 2011.
- [7] G. Linden, B. Smith and J. York, Amazon.com recommendation: Item-to-item collaborative filtering, *IEEE Internet Compute* **7**(1) (2003), 76–80.
- [8] S.S. Manvi, N. Nalini and L.B. Bhajantri, Recommender system in ubiquitous commerce, in: *International Conference on Electronics Computer Technology*, IEEE, 2011, pp. 434–438.
- [9] D. McSherry, Explaining the pros and cons of conclusions in CBR, in: *European Conference on Case-based Reasoning*, Springer Berlin Heidelberg, 2004, pp. 317–30.
- [10] *Transactions on Knowledge and Data Engineering* **17**(6) (2005), 734–749.
- [11] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes and M. Sartin, Combining content-based and collaborative filters in an online news paper, in: *Proceedings of ACM SIGIR Workshop on Recommender Systems: Algorithms and Evaluation*, Berkeley, CA, 1999.
- [12] M. Balabanovic and Y. Shoham, Fab: Content-based, collaborative recommendation, *Communications of the ACM* **40**(3) (1997), 66–72.
- [13] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, Item based collaborative filtering recommendation algorithm, in: *Proc 10th International Conference on World Wide Web*, 2001, pp. 285–295.
- [14] I.H. Witten, E. Frank and M.A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, Elsevier; 2011.
- [15] D.J. Hand, H. Mannila and P. Smyth, *Principles of Data Mining*, MIT press; 2001.
- [16] Y. Zhao, *R and Data Mining: Examples and Case Studies*, Academic Press; 2012.
- [17] R. Jensi and G. Wiselin Jiji, An improved krill herd algorithm with global exploration capability for solving numerical function optimization problems and its application to data clustering, *Applied Soft Computing* (43) (2016), 248–261.