Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 171–179, DOI 10.12694/scpe.v23i4.2015

INTEGRATING COLLABORATIVE FILTERING TECHNIQUE USING RATING APPROACH TO ASCERTAIN SIMILARITY BETWEEN THE USERS

PAVITHRA C*AND SARADHA M†

Abstract. The recommender system handles the plethora of data by filtering the most crucial information based on the dataset provided by a user and other criterion that are taken into account.(i.e., user's choice and interest). It determines whether a user and an item are compatible and then assumes that they are similar in order to make recommendations. Recommendation system uses Singular value decomposition method as collaborative filtering technique. The objective of this research paper is to propose the recommendation system that has an ability to recommend products to users based on ratings. We collect essential information like ratings given by the users from e-commerce that are required for recommendation, Initially the dataset that are gathered are sparse dataset, cosine similarity is used to find the similarity between the users. Subsequently, we collect non-sparse data and use Euclidian distance and Manhattan distance method to measure the distance between users and the graph is plotted, this ensures the similar liking and preferences between them. This method of making recommendations are more reliable and attainable.

Key words: Collaborative filtering; Euclidian distance method; Manhattan distance method; SVD.

AMS subject classifications. 68T35

1. Introduction.

Singular value decomposition is a significant technique used for recommendation system. SVD is widely used in developing various models. In precise SVD is used in e-commerce recommendations, it contributes to decrease the range of the data sets values by using matrix factorizing method, (cuts the space dimensions from N-Dimensions to K-Dimensions, where strictly K is lesser than N). Since singular value decomposition is a method of linear algebra, it uses the matrix structure to solve the system of problems where each column represents an item and each row user represents the user. Each element present in the matrix are the ratings that are given by the users to the respective items.

1.1. Singular Value Decomposition.

Matrix Factorizing is done using singular value decomposition method and given by

$$A = PSV^T$$

The dataset value (user-item-ratings) of the matrix are decomposed using singular value decomposition method. This helps the matrix decompose into three types as given below, thus the factors of the matrix are obtained,

$$A_{m,n} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{pmatrix} \tag{1.1}$$

*Research Scholar, Department of Mathematics, Reva University, Rukmini knowledge park, Bengaluru 560064, India (pavithracshekar5@gmail.com)

†Associate Professor, Department of Mathematics, Reva University, Rukmini knowledge park, Bengaluru 560064, India (ishuharisri@gmail.com)

$$A_{m,n} = \begin{pmatrix} p_{1,1} & \cdots & p_{1,r} \\ \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,r} \end{pmatrix} \begin{pmatrix} s_{1,1} & \cdots & 0 \\ 0 & \vdots & \vdots \\ \vdots & \vdots & s_{r,r} \end{pmatrix} \begin{pmatrix} v_{1,1} & \cdots & v_{1,n} \\ \vdots & \vdots & \vdots \\ v_{r,1} & \vdots & v_{r,n} \end{pmatrix} \tag{1.2}$$

where A is a $m \times n$ utility matrix, P is a $m \times r$ left singular orthogonal matrix, this represents the affinity between users and latent factors. Latent factors are defined as describing the nature of the items. S is a $r \times r$ diagonal matrix, recounts the strength of each latent factor and V is a $r \times n$ diagonal right singular matrix, which indicates the relationship between items and latent factors. From (1.1) and (1.2) we see that, the utility matrix is decomposed into orthogonal matrix and diagonal matrix.

Latent factor contributes to reduce the dimension of the matrix in singular value decomposition method. The below equation is the mapping that accelerates a clear representation of relationships between users and items.

$$r_{u,i} = x^t \cdot y_u$$

Let xi represent each item, and yu represent each user. The expected rating by a user on an item $r_{u,i}$ can be given as Here, $r_{u,i}$ is a form of factorization in singular value decomposition.

1.2. Recommendation system. The role of recommendation system is to give suitable and relevant suggestions to the users, this determines the relationship between the users and the products, and also helps the organization to provide an appropriate product suggestion. Recommendation system is incorporated with collaborative filtering method, which uses the concepts of singular value decomposition. Most of the e-commerce uses recommendation system for their customer support interface.

Recommendation system has few beneficial aspects, to both user and organization. It contributes to avoid the local transaction caused by choosing products or items. Therefore, It helps in getting the suitable suggestion for the relevant search depending on the previous search history.

Systems that make recommendations employ a variety of technologies. These systems can be divided into two main categories called content based filtering and Collaborative filtering.

1.3. Content-based Filtering. Making decisions using similar features is a machine learning technique and are called content based filtering. This method is frequently applied in recommender systems, that are algorithms created to promote or suggest products to users based on data gathered about the user.

1.4. Collaborative Filternig. Collaborative Filtering claims an assumption, this approach describes the user who liked a product now will tend to have the similar likings in the future. This is done by analyzing the behavior of the user in the previous searches, likes, add carts, ratings etc. by using this approach the model finds the relationship and inter-dependence of two variables (i.e., users and items). Thus, Collaborative filtering is performed by using the technique of singular value decomposition.

1.5. Rating-based Recommendation. The familiarity of a product or Item can depend on high user ratings. In recommender system the customers or users give their perspective in the form of ratings and comments also users tend to give “explicit feedback”, Also further, clicks, shop for, and search history are called “implicit feedback”.

Depending on the ratings given to the products or the items the product suggestions are done to the users, in advancement to the same depending on the previous searches and choices them recommendation is done. This helps in filtering the information and preferences achieved by the item given by the user.

2. Literature Review. In [1] by Utkarsh Gupta et al. Hierarchical clustering technique was used for product recommendation system, this approach gives the results of low error and better clustering of similar items. In this research paper voting technique was used to determine the rating values of the product.

In [2] by Xiaoyuan Su and Taghi M. Khoshgoftaar. discuss about the challenges and the main charges of collaborative filtering, such as privacy protection, data sparsity, scalability etc., this paper also describes three different types of collaborative filtering technique, like model based CF which finds the relationship between the user and the items. Memory based and hybrid CF, it is the combination of both the methods model based and memory based collaborative filtering technique. This was the survey based research paper.

In [3] by Harpreet Kaur et al. presents the research work in the hybrid system, which involves the combination of conceptual contents and the CF algorithms. This means that the recommendations are done based on the stuffs and concepts involved in any closure work. It basically progresses on user with user and also user with items for the recommendation.

In [4] by Geetha G, Safa M, Fancy C, Saranya D. formulated for movie recommendation system, which helps in recommending the movies to the users. This revises the existing movie data base set for few relevant information like ratings of the movie, genre of the movie and popularity. This paper is the combination of content based collaborative filtering and hybrid filtering method. It helps in achieving more precise results for recommendations.

In [5] by Sarwar B, Karypis G, Konstan J, Riedl J. used CF recommender system that collects ratings from users for products in a particular domain. Association inferences, which are exceedingly time- and scalability-intensive and have a very likely temporal complexity, were also a reliance on CF algorithms used for recommendation systems. The more efficient and scalable matrix operations are used in modern methods.

In [6] by C. S. M. Wu, D. Garg, and U. Bhandari. have used collaborative filtering technique for suggesting in e-commerce. Here the suggestions are done based on the user ratings. This paper has included Apache Mahout Framework, and the efficiency are checked between user based and item based recommendations.

In [7] by Pennock and Horvitz. contributed to the clustering method of solving the problem, it works on different data set size of the user feature vectors. A new concept of LGS was introduced to check the perform ability of the collaborative filtering methods. LGS stands for Latent Genre Space.

In [8] by Debadrita Roy, Arnab Kundu. have contributed to the recommendations with respect to movies. This is achieved with the help of collaborative filtering technique. Expectation Maximization Algorithm are used for movies recommendations. Clustering the data was also used for recommendation system desings.

In [9] by J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. developed a research paper which is the open ended research paper, where the author ends the paper with the question and the challenged faced by the recommendations like privacy policies and terms conditions. Design decisions are developed for rich interaction interface. The primary use was to privacy structuring regarding rating systems.

In [10] by Desrosiers C, Karypis G. Implemented the simple logic, the product used or liked by the user in the past will also be liked in the future. Here collaborative filtering techniques are used and developed to meet the expectations of the users.

3. Background of the work.

3.1. Collaborative Filtering. The most common method for creating recommendation systems is collaborative filtering, which has been effectively used in numerous applications. The CF recommender system gathers user feedback for products in a certain domain in the form of ratings. Also CF algorithms which was for recommendation systems relied on association inferences, which are extremely time- and scalability-intensive and have a very probably about time complexity. Modern techniques use matrix operations, which are more effective and scalable [5].

3.2. Memory based collaborative filtering. It is based on both the item's description and the user's preference profile. In memory-based collaborative filtering, we use key-terms in addition to the user's profile to propose items by indicating the user's preferred likes and dislikes. In other words, products that were previously favored are recommended via a memory-based CF algorithm. It looks at previously rated things and suggests the best option [3][5].

3.3. User-based prediction system. When compared, User-based filtering method it is expected to be better to be working with huge amounts of data, while item based collaborative filtering methods are used when the data set are small in size.

We say that user based collaborative filtering technique is used for the large amount of data. As discussed above CF is used in determining the likings and disliking of the product or the items. Here Figure 3.1 represents the flow chart for user-based prediction system. This is done by monitoring the previous search history of

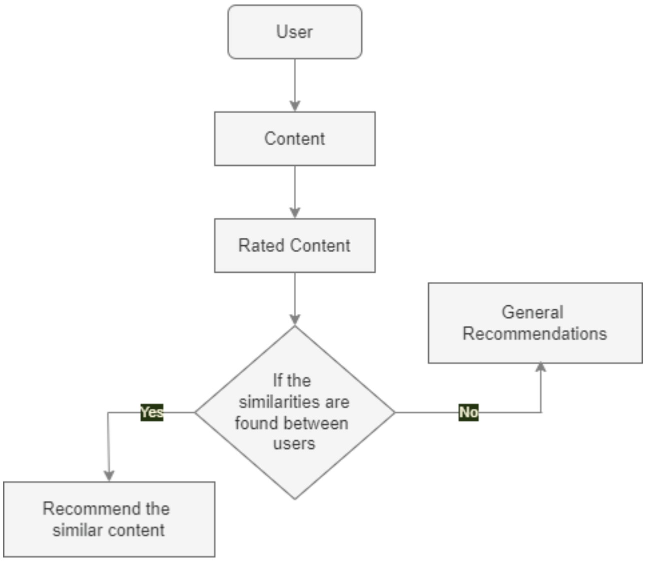

User
Content
Rated Content
If the similarities are found between users
Yes
No
General Recommendations
Recommend the similar content

Fig. 3.1: Flow Chart

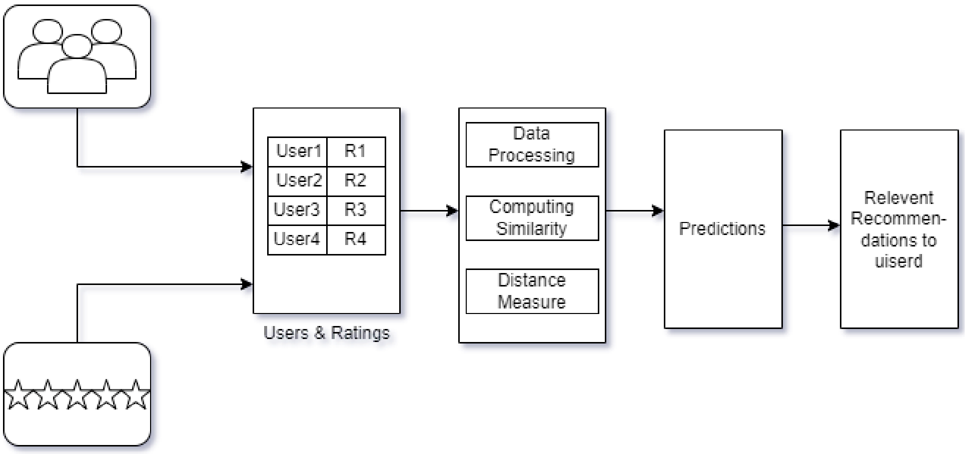


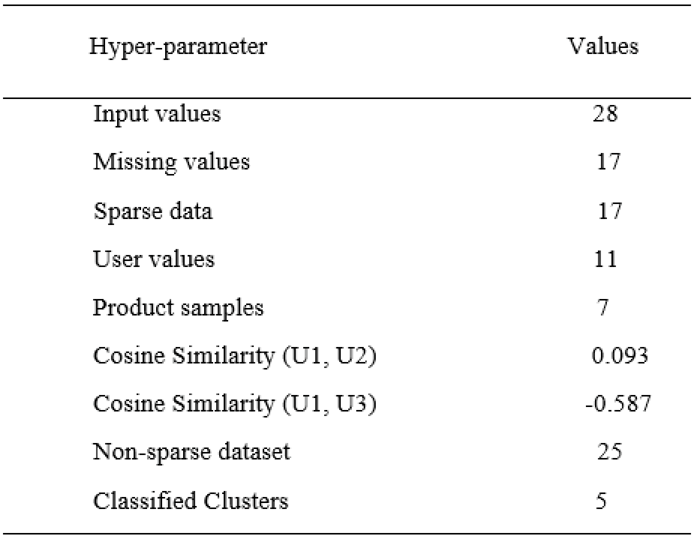
Fig. 4.1: Proposed Structure

any users. Today most of the e-commerce are using the Collaborative filtering technique for improving the recommendation suggestion to their customers and users.

4. Proposed System and Methodology. To find the similarity between the users, the following claim is raised,tabular column(users vs items)the User1 who have rated 4.5/5 for “Item1” and “Item3” and if User2 has rated 4/5 for product “Item1” and “Item2”,then we conclude that user1 and user2 have similar liking. Also, “Item1” will be recommended to User1 and “Item3” will be recommended to User 2.

The structure of the proposed system is given in Figure 4.1. The set of data(ratings) given by the users are collected from e-commerce and tabulated. The obtained data are pre-processed and the missing values are computed. The similar users are identified using cosine similarity and centered similarity. An illustration was performed by collected twenty-five data samples from e-commerce, graph was plotted across users and their ratings for respective products. The distance between the sample points was computed and classified into

Table 4.1: Hyper-parameter values



Hyper-parameter	Values
Input values	28
Missing values	17
Sparse data	17
User values	11
Product samples	7
Cosine Similarity (U1, U2)	0.093
Cosine Similarity (U1, U3)	-0.587
Non-sparse dataset	25
Classified Clusters	5

clusters, and they represent the users with similar likings and preferences.

In Specific we are using cosine similarity method and distance measure i.e., Euclidean distance method and Manhattan distance method to find the similarity between the users. Here Cosine similarity finds the comparability between two vectors in inner product space establishes whether two vectors are roughly pointing in the same direction by calculating the cosine of the angle between them. In text analysis, it is frequently used to gauge document similarity. The method of finding the distance between any two points is known as Euclidean distance method. It helps in determining the length of the line segment between any two points. Also Manhattan method is used as a metric for measuring distance between any two points in an N-dimensional vector space.

This new approach of identifying the similarity and common likings between the users was performed because the gathered data was sparse data, making it difficult to make the right suggestions to the users. As a result, we start by turning all of the sparse data into easily recognizable unique values, then we continue to compute how similar they are to one another, using cosine similarity method. Subsequently, we collect a non-sparse dataset from e-commerce to compute similarity between the users, we utilize the distance measure approach to calculate similar users with particular items.

5. Implementation. The role of the proposed system is to determine the missing rating of the products or items. The tabular column of the rating data sets is computed by initially assigning nil values and further the values are neutralized. Cosine similarity method is implemented between the users to find the similarity between them and it is given by

$$Cosinesimilarity(u,v) = Similarity(u,v) = \frac{\sum(r_{u,i}-r)(r_{v,i}-r)}{\sqrt{\sum(r_{u,i}-r)^2}\sqrt{\sum(r_{v,i}-r)^2}} \tag{5.1}$$

Table 5.1 represents the score spread of ratings given by four different users for seven different items or products. Since not all the users rate all the seven products, there exists few missing values in the datasets. For further analysis, we require to complete the table by marking the missing values with zero and converting the missing dataset values to sparse data. A dataset with sparse data is one in which a sizable portion of the cells do not really contain any data, but occupies the storage space in the file.

By using (5.1), similarity betweeen user1 and user2 is given by Sim(U1, U2) = 0.2 and Sim(U1, U3) = 0.5. This is contradicting the theory of similarity for common liking recommendation, and thus, the following table

Table 5.1: User Ratings

	user1	**user2**	**user3**	**user4**	**user5**	**user6**	**user7**
Item1	4			5	1		
Item2	5	5	4				
Item3				2	4	5	
Item4		3					3

Table 5.2: Generalizing missing values

	user1	**user2**	**user3**	**user4**	**user5**	**user6**	**user7**
Item1	4	0	0	5	1	0	0
Item2	5	5	4	0	0	0	0
Item3	0	0	0	2	4	5	0
Item4	0	3	0	0	0	0	3

Table 5.3: Cosine similarity

	user1	**user2**	**user3**	**user4**	**user5**	**user6**	**user7**
Item1	4	0	0	5	1	0	0
Item2	5	5	4	0	0	0	0
Item3	0	0	0	2	4	5	0
Item4	0	3	0	0	0	0	3

Table 5.4: Centered Values

	user1	**user2**	**user3**	**user4**	**user5**	**user6**	**user7**
Item1	2/3			5/3	-7/3		
Item2	1/3	1/3	-2/3				
Item3				-5/3	1/3	4/3	
Item4		0					0

was evolved, sim(A,B)= cos(rA,rB). In this method rating values are Ignored. This is contradicting the theory of similarity for common liking recommendation, and thus, the table 5.2 was evolved.

Similarity(U1, U2) = 0.38 and Similarity(U1, U3) = 0.322. Therefore, Sim(U1, U2) is greater than Sim(U1, U3). The difference between them are not significant. The similarity is not precise and accurate.

By Neutralizing the Table 5.3 we get Table 5.4.

By using (5.1) we compute Similarity(U1, U2)= 0.093 and Similarity(U1, U3)= -0.587. Table 5.4 represents the centered values of ratings. Therefore, Sim(U1, U2)is greater than Sim(U1, U3). This shows that there exists a similarity between User1 and User2 and non-similarity between User1 and User2 (Since, Negative value).

Our next step, is to use distance measure methods which makes conclusion about an item, this method will calculate the “distance” between the target movie and every other item in its database, then it ranks its distances as the most similar recommendations.

5.1. Distance Measure. Here we are using methods for calculating distance between points are Euclidian and Manhattan distance are the two methods that are used for computing the relationship between any two variables and objects. Euclidean distance method is the familiar or called basic method of finding the distance between the points.

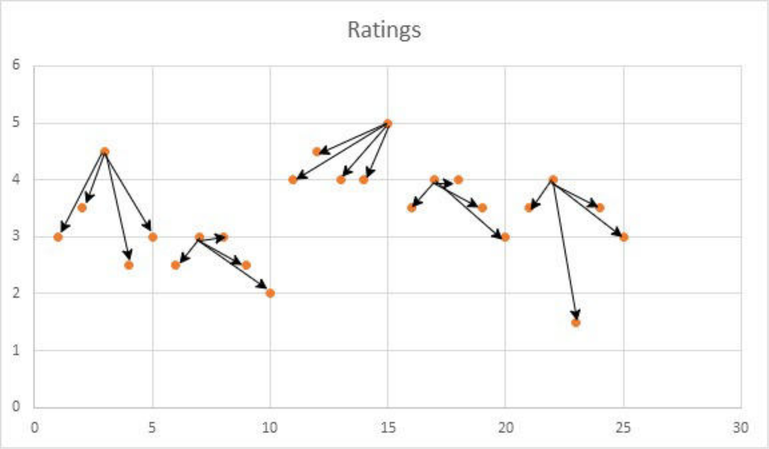


Fig. 6.1: Scatter Plots

Euclidean Distance: Euclidean distance is defined as, calculating the square root of the sum of the squared differences between a new point (x) and an existing point (y).

$$d_(i,j) = \sqrt{(x_{i,1} - x_{j,1})^2 + (x_{i,2} - x_{j,2})^2 + ... + (x_{i,3} - x_{j,3})^2} \tag{5.2}$$

or

$$d_(i,j) = \sqrt{\sum (x_{i,k} - x_{j,k})^2} \tag{5.3}$$

Manhattan Distance: Manhattan distance is defined as finding the absolute values of the difference of real vectors, and it is given by

$$d_(i,j) = |(x_{i,1} - x_{j,1})| + |(x_{i,2} - x_{j,2})| + ... + |(x_{i,n} - x_{j,n})| \tag{5.4}$$

or

$$d_(i,j) = \sum |(x_{i,k} - x_{j,k})| \tag{5.5}$$

6. Experimental Setup. We have considered 25 users and the corresponding rating values from the E - Commerce, the chosen values are plotted in scattered graph shown in the below Figure 6.1 and they are classified using clustering technique based on the nearest neighbor concept. Clustering is defined as making groups with the existing data, to obtain a new form of clusters or groups.

This results in achieving the similarity of the data in same group and dissimilarities when compared to the data values from one grouped values to another. Often, distance measures are used, an appropriate arbitrary point was chosen from cluster and two methods of calculating the distance between every other point from an arbitrary point (i.e., Euclidean distance method and Manhattan distance method) was performed, this is achieved by using multiple iterations, and the multiple bar graph are plotted. In this experiment we use (4) Euclidean distance method and (6) Manhattan distance method to find the distance between them.

The proposed system is contributing in classifying the similar types of the users based on the rating values. Table 6.1 represents the resulting values of distance measure. The proposed system is developed using collaborative filtering approach which helps e-commerce to find the similar likings of the users based on their reviews and reactions, and the appropriate suggestion are filtered. Subsequently, we find the subset of the group of the users from the bigger data sets that are collected from the e-commerce.The above Figure 6.2 represents the similarity between clusters and ratings.

Table 6.1: Comparison values

	Euclidean Distance K-Means	**Manhattan Distance K-Means**
c1	3.08	5
c2	1.22	2
c3	1.87	3.5
c4	1.22	2
c5	2.78	4.5

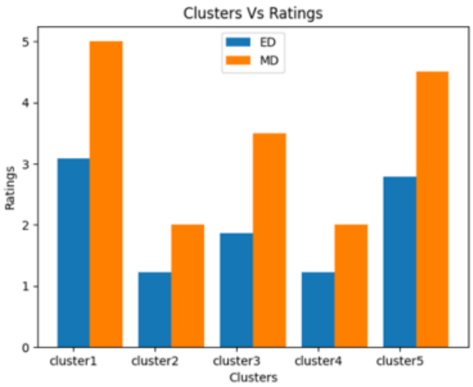


Fig. 6.2: Cluster Vs Rating

Here the datasets which was gathered from e-commerce is the set of sparse data; thus this does not support in achieving the appropriate recommendations to the users. Therefore, as the first step we convert all the sparse data into a distinguishable unique values and continue to calculate the similarity between them, and to measure the strength of the similar users we use distance measure method to determine the correlation between the users and the products.

This approach of finding the similarity between the users is a reliable method if the datasets that are obtained are sparse data. The methods like cosine similarity and distance measure methods which are implemented on the sparse data produced desired results. This method of making recommendations are more reliable and attainable for sparse data samples. This proposed system works well for sparse data and produces the reliable results.

7. Limitations of the work proposed and their future improvement.

7.1. Limitations. For this proposed system, the data which is gathered from the e-commerce is more often the sparse dataset. Thus, determine the missing value and averaging the datasets become the mandatory steps to be carried out to find the similarity between the users. In this research paper the recommendations are done only based on the ratings given by the users and neglecting other parameters.

7.2. Future Work.

a. For the aim of making recommendations, we'll strive to combine the genre dimension with other product dimensions (reviews, reactions, etc.). It will improve how effective our suggestion is.
b. Expand the user and product databases respective data sizes.
c. Use additional collaborative filtering methods to compare recommendations.

8. Conclusion. In this research paper, we have developed a recommendation system for e-commerce based on rating system. We used singular value decomposition method as collaborative filtering technique.

This approach helps us in identifying the similarities between the user and make relevant suggestions of the product. We gather vital data, such as user evaluations provided by e-commerce, which are necessary for recommendations. Cosine similarity is utilized to determine how similar the users are to one another for sparse data. For the non-sparse dataset, we assess the distance between users and depict the graph using the Euclidian and Manhattan distance methods, which guarantees that their tastes and preferences are similar. This approach to giving recommendations is more trustworthy and practical.

REFERENCES

[1] Utkarsh Gupta, Dr Nagamma Patil, "Recommender System Based on Hierarchical Clustering Algorithm Chameleon" IEEE International Advance Computing Conference (IACC) 2015.

[2] Xiaoyuan Su and Taghi M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques Advances in Artificial Intelligence Volume 2009, Article ID 421425, 19 pages, October 2009.

[3] Harpreet Kaur Virk, Er. Maninder Singh," Analysis and Design of Hybrid Online Movie Recommender System "International Journal of Innovations in Engineering and Technology (IJIET)Volume 5 Issue 2,April 2015.

[4] Geetha G, Safa M , Fancy C, Saranya D4, "A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System" National Conference on Mathematical Techniques and its Applications (NCMTA 18) 2018.

[5] Sarwar B., Karypis G., Konstan J., Riedl J., "Item-based Collaborative Filtering Recommendation Algorithms," Published in the Proceedings of the 10th international conference on World Wide Web, Hong Kong, May 15, 2001.

[6] C. S. M. Wu, D. Garg, and U. Bhandary, "Movie Recommendation System Using Collaborative Filtering," In 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), pp. 11-15, IEEE, 2018 November.

[7] D. M. Pennock and E. Horvitz, Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach, In IJCAI Workshop on Machine Learning for Information Filtering, Stockholm, Sweden, August 1999.

[8] Debadrita Roy, Arnab Kundu, "Design of Movie Recommendation System by Means of Collaborative Filtering" International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 4, April 2013.

[9] J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. "Collaborative Filtering Recommender Systems" The Adaptive Web, LNCS 4321, pp. 291 – 324, Springer-Verlag Berlin Heidelberg 2007.

[10] Desrosiers C, Karypis G, "A comprehensive survey of neighborhood-based recommendation methods". In: Ricci F, Rokach L, Shapira B, Kantor P, Recommender systems. Springer, Boston, pp 107–144, 2011.

Edited by: Vinoth Kumar
Received: Jun 11, 2022
Accepted: Oct 10, 2022