Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 7 (2017) pp. 1969-1979

http://www.ripublication.com

DECORS: A Simple and Efficient Demographic Collaborative Recommender System for Movie Recommendation

M.Sridevi
Anurag Group of Institutions, Hyderabad, Telangana, India.

Dr .R.Rajeswara Rao
JNTU Kakinada, Vizianagaram, Andhrapradesh, India.

Abstract

With the explosion of internet users on the web, finding the right choice for the user is a tedious task. Collaborative filtering one of the most commonly used Recommender system recommends the movies based on user preferences. However, it requires expensive computations to find the similarity as the number of users and movies increases. Hence, DECORS a Movie Recommender is proposed which provide a solution for the aforementioned problem. DECORS is based on collaborative filtering which initially partitions the users based on demographic attributes, then using k-means clustering algorithm clusters the partitioned users based on user rating matrix. It reduces the expensive computations to identify similar users in order to predict movies when compared with traditional collaborative filtering approach. The proposed system also sorts the movies in the increasing order of user's preferences. The proposed framework is assessed by using the performance measurement MAE(Mean Absolute Error). The results proved that proposed system is more efficient when compared against traditional methods.

Keywords: Collaborative filtering, Demographic, clustering.

I. INTRODUCTION

With an overwhelming growth of information on the Internet in recent years, there are various web portals offered by different Ecommerce web sites. As the number of online users and services are increasing, an efficient technique need to be introduced to manage large amount of data. Finding the right choice for the online user is a tedious task with an information overload. To cope up with information overload problem, Recommender systems [1] have been implemented in various E-commerce websites to guide Internet users. The key challenges posed with the tremendous growth of internet users and products are generating high quality recommendations and performing recommendations faster. However these techniques fail to provide efficient recommendations. Therefore there is a need to optimize the existing system.

Recommender system can be classified as Collaborative filtering, Content based filtering, Demographic filtering and Hybrid Recommenders system. The most traditionally and widely used Recommender system is Collaborative filtering [2]. It assumes the users who have similar preferences in the past have similar preferences in the future. It recommends to the users based on the preferences of other users who have same tastes and likings. Content based Recommender System [3] recommends items to the user similar to those items preferred by the user in the past. Demographic Recommender system recommends items based on demographic information of the users [4]. It does not require users ratings or knowledge of the item and thus can overcome cold start problem.

Collaborative filtering is one of the successful Recommender system which generates recommendations based on similar user preferences. However, as the number of users and products increases, it resulted in the failure of real time requirement. It requires expensive computations to compute the similarity as the number of users and movies increases in the database. Thus, there is a need for a system which can produce high quality recommendations quickly when there is an increase in the cardinality of user data base.

To overcome the problems of cardinality and efficiency, an efficient Collaborative filtering Recommender DECORS is built using Divide and conquer approach[8]. The system combines both Collaborative filtering technique and user Demographic features to recommend the movies user interested in.

The proposed approach generates personalized recommendations by employing collaborative filtering on user demographic data. Initially, it performs preprocessing based on user demographic features thereby, reducing the search space and subsequent similarity computation[11]. One major advantage of preprocessing user's data is that working with smaller data which contains the entire database essence is efficient and speed. Sufficient data reveal important facts. The time consumed to compute similarity from the processed data is less than finding similarity from complete data.

It is assumed here, that users with similar demographics will ease in identifying the users with similar preferences and tastes[5]. It not only exploits individual user features but also user ratings. The system satisfies large amount of users and customers in growing database.

II. LITERATURE SURVEY

1. Collaborative Filtering

Collaborative filtering, the traditional Recommender system is based on rating structure usually represented as User-Movie Rating matrix. Each cell value represents the rating of a movie by the user. It predicts the ratings based on similarity measures like Pearson correlation coefficient, cosine similarity, Euclidean distance measure, etc., Collaborative filtering can be classified into two types : Memory based CF and Model based CF. Memory based CF predicts the ratings using the entire user-item database of users who are similar to the active user whereas Model based CF predicts the ratings by using the constructed model. Collaborative filtering often suffers from several issues which include :

Sparsity: Most often users do not rate the movies which results in sparsity of data.

Cold start: To recommend a new item or for new user who has not yet rated any movies, it is very difficult as there exists no user information.

Scalability: To handle millions of users and movies over Internet, CF computations to find similar users grow exponentially and becomes expensive.

2. Content Based Recommender System.

Content based recommendations are based on the user individual preferences and tastes. It recommends the movies preferred by the user in the past.

Content-based Recommender system often suffer from the following issues:

Limited content analysis It is difficult to recommend if there is a limited content about the user profile

Overspecialization restricts users to items similar to the ones defined in their respective profiles and thus new items and other options are not discovered.

3. Demographic Recommender system.

Demographic Recommender system generate recommendations based on the user demographic attributes. It categorize the users based on their attributes and recommends the movies by utilizing their demographic data [4]. In contrast to collaborative filtering and content based recommender system, it is easy to implement and does not require user ratings.

4. Hybrid Recommender system.

Hybridization of demographic and collaborative filtering approaches had been employed to solve cold start problem [10]. Hybrid model based approach has been applied on Movie data set to enhance recommendation quality. Additionally Collaborative filtering and Demographic based approach had been used to modify similarity calculation[1][2].

In contrast, this paper proposes a novel approach to enhance the Recommendation quality by utilizing user demographic data provided by the users explicitly and Collaborative filtering approach based on user ratings on movies.

III. METHODOLOGY

The proposed framework consists of 3 phases: Data Source, Similarity Computation phase and Recommendation phase (as shown in figure 1).

Data input phase collects users demographic data and ratings of the user which are the two main sources of input. Similarity computation phase uses demographic data of users and partitions into groups having same demographic features. These partitions are subsequently used to calculate user similarity based on the ratings given by the users on movies, thus forming another sub cluster of similar users [7] by employing efficient K-means clustering algorithm [6]. Each user is likely to belong to one of 'n' clusters and these clusters contain similar users than those in other clusters yielding more accuracy. Finally Recommendation phase recommends movies to the target user based on the opinions of similar users i.e., on which cluster the user belongs, which is computed from user demographic data collected from user profile and user ratings. An advantage of the proposed system is that it reduces the time to generate recommendations by creating user subsets.

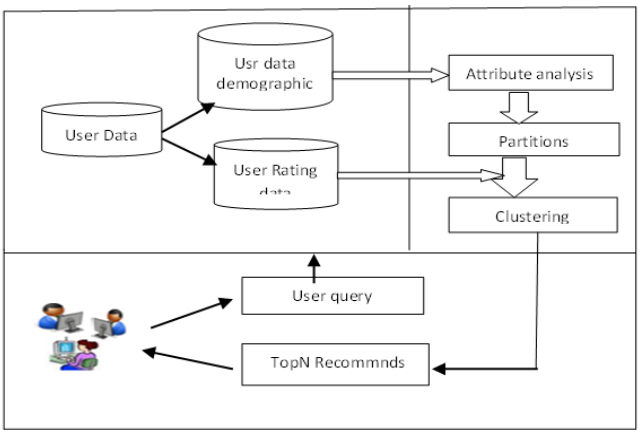

Figure.1. Framework of the Proposed system

Euclidean distance measure is used to calculate the distance between the users to find similar users based on rating matrix and using K-means algorithm, forming the cluster of similar users.

The distance between users x^1 and x^2 is calculated below:

$$r_2(x^1,x^2) = \mathrm{Dist}(x^1,x^2) = \sqrt{(x_1^1 - x_1^2)^2 + (x_2^1 - x_2^2)^2 + \cdots + (x_D^1 - x_D^2)^2}$$
$$= \sqrt{\sum_{d=1}^{D}(x_d^1 - x_d^2)^2} \quad (1)$$

We compute the similarity between two users x and y by performing the following calculations

$$\mathrm{sim}(x,y) = \frac{1}{1+r_2(x,y)} \quad (2)$$

We recommend movies to user by weighting the ratings of each "critics" by the similarity to user. An overall score for each movie is obtained by summing these weighted scores. If u is the user and we have C critics, then the estimated score given to a movie z by user u, $sc_u(z)$ is obtained as follows:

$$sc_u(z) = \frac{1}{\sum_{c=1}^{C} sim(x^u,x^c)} \sum_{c=1}^{C} sim(x^u,x^c) \cdot sc_c(z) \quad (3)$$

Proposed Algorithm:

Input: *No.of clusters K, demographic data of users, user-item rating matrix;*

Output *: Top n Recommendations to the target user.*

Begin:

1. *Consider User set= $\{U_1,U_2,U_3, \ldots. U_m\}$*
2. *Consider Movie set=$\{M_1,M_2,M_3 \ldots.M_n\}$*
3. *Consider Rating Matrix U_m X M_n.*
4. *Extract any two attributes from user set U_m. Let it be Age group and Gender .*
5. *Partition users by considering all possible combination values of selected attributes i.e{M-M},{M-F},{S-M},{S-F}. Let it be PU=$\{PU_1,PU_2,PU_3, PU_4\}$ where every user may belong to any one of the partition PU_i.*
6. ***for*** *every PU_i where i=1 to 4*
   i. *Select 'k' users randomly as the initial cluster centre CU=$\{CU_1,CU_2, \ldots CU_k\}$*

ii. For each user Um ∈ PUi
a.For each cluster center CUj ∈ CU
b.Calculate sim(Um, CUj) based on user-movie rating.

iii. Update the clustering centers.

iv. Repeat step 6.ii until clustering centers doesn't change.

7. ***end for**.*
8. *Allocate Active user U_a to the most similar cluster generated from partition lists.*
9. *Generate the predictions and recommend the top movies for the target user.*
10. ***End***

IV. EXPERIMENTAL SETUP

In this section, simulation of the Proposed system, Software interfaces, datasets, database tables, evaluation standards and experimental results are discussed.

1. **Simulation of DECORS.**

DECORS, a Movie Recommender is a web application which is built using Bootstrap, a responsive mobile first framework. A user logs on to the system and has to register in the Registration page. This page is provided with a set of attributes for which the user has to input the data. After registration he/she has the option of browsing the list of movies, searching for movie information or can reach the recommendation page where ratings are given for different movies and a custom taste profile is built. The system learns continuously the user preferences, and keeps getting better with usage over time.

To recommend the movies accurately and efficiently k-means clustering algorithm is used by pre filtering the data base on user features. Demographic information i.e., gender and age group of a user partitions the users. Each of the categorical attribute has 2 possible values – {Male, Female} and {Youth (15-24), Adult(25-64) }thus resulting in combination of 4 partitions. The cardinality of the users data has a significant impact in computing the similarity of users.

2. Software Interfaces:

We have experimented and implemented a prototype for Recommender system.

We use HTML, CSS for the design of the application, JavaScript and JSP for implementing the algorithm, MySQL is used for creating the database, Apache Tomcat is used as a server. The operating System used while developing the application is Windows 8.

3. Database Design

The information required for our proposed system has been taken by extensively searching free online movie databases. It consists of 1060 ratings, and is evaluated by 210 users, on 90 movies. Every user rated around 8 movies with range of 1(bad) to 5(Excellent).

The format of user movie rating matrix is depicted in table 1 where each entry represents the rating of a movie by the user and table2 demonstrates sample demographic data of few users

Table 1. User rating matrix

	m_1	m_2	m_3	m_4	m_5
u_1	3	4	4	5	5
u_2	4	5	5	3	4
u_3	3	5	5	1	4
u_4	5	3	4	5	3
u_5	5	4	4	4	5

Table 2. Sample user demographic data

Uid	Name	Gender	Age Group	Occupation
u_1	Beckham	M	A	Doctor
u_2	Jose	F	Y	Educator
u_3	Robin	M	A	Engineer
u_4	Jenni	F	A	Lawyer
u_5	Jacob	M	Y	Doctor
u_6	Cherry	F	Y	Engineer

which includes user id, name, gender, age group and location. The data pertaining to the movie i.e., name, description, trailer link etc., is depicted in table 3.

Table 3. Sample movie data

MID	Mov Name	url	Details	Genre
200	Captain America	img/civilwar2.jpg	Political pressure	Action
201	Titanic	img/titanic.jpg	Seventeen years	Romantic
202	The Martian	img/themartian.jpg	Whenastronauts	Science fiction
203	Star wars	img/theforce.jpg	Thirty years	Science fiction
204	Stevejob	img/steve.jpg	With public	Documentary
205	The Man …	img/man.jpg	The Man	Documentary

4. Evaluation standards

The rating data is divided into two sets which comprises of 80% training data and 20% of test data. To validate the efficiency of Collaborative Filtering Recommender system, researchers have proposed many evaluation standards. This paper considers Mean Absolute Error as the standard measure to assess the accuracy of the system.

$$\text{MAE}=\frac{1}{n}\sum_{i=1}^{n}|p_i - r_i| \tag{4}$$

where p_i and r_i is prediction ratings and actual ratings for movie ‘i’, and n is number of rated movies. The lower the MAE, the more accurate the prediction would be, providing better quality recommendations.

5. Experimental Results

Our experiments proved that the proposed system DECORS generates efficient and accurate predictions when compared to user based collaborative filtering. Similarity computation of users is not required for all pairs of users. The table 4 shows that the performance of the proposed system has been improved than the traditional system. We compare the results between traditional method and proposed method and is represented in figure 2. It shows that the proposed system is far better than existing one.

Table 4. MAE for proposed system and Existing system.

	No.of Users	Traditional CF	Proposed system
MAE	10	0.89	0.68
	50	0.65	0.49
	90	0.32	0.23
	130	0.18	0.15
	170	0.08	0.07
	210	0.06	0.05

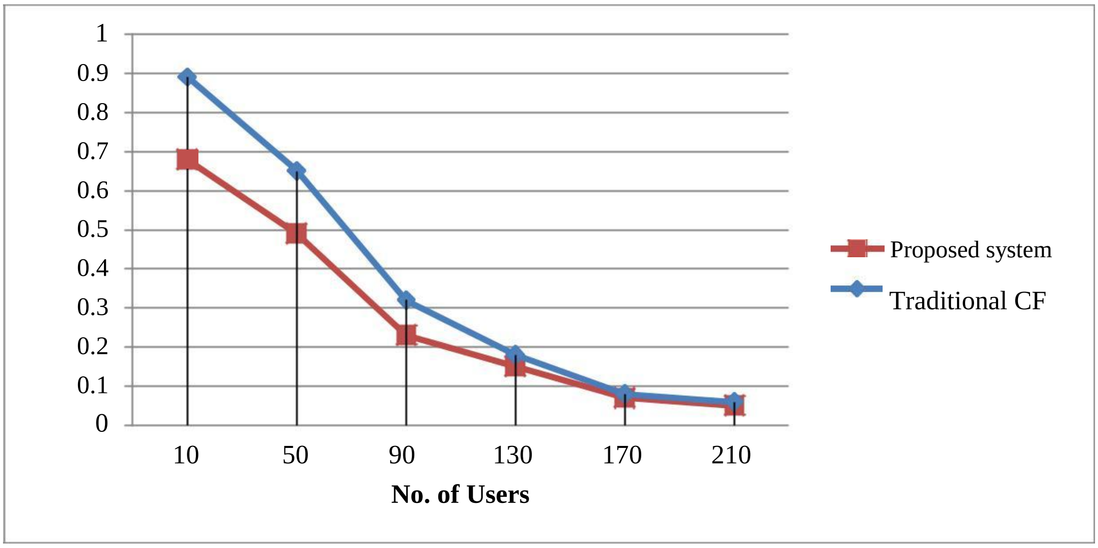

Figure 2. MAE for proposed system and Existing system

V. CONCLUSION

In this paper, we have proposed DECORS, a Movie Recommender which recommends movies based on the user demographic attributes and user ratings. A novel approach is presented to solve the problems of cardinality and efficiency. In this approach, users are partitioned based on user demographic features and each user partition is again clustered based on users ratings on movies. In this paper, K- means clustering algorithm is used to cluster the users. Based on targets users' demographic attributes and user ratings, similar users are identified and the recommendations are generated. The proposed system joining user demographic features and user ratings is more accurate and more scalable than the traditional collaborative filtering. Evaluation is carried on a small set of users with different attributes, and the result proved to be positive in terms of accuracy and speed. In future we would like to implement on a larger set of users and compare the results with traditional methods.

REFERENCES

[1] Kantor, P. B., Rokach, L., Ricci, F., & Shapira, B. (2011). Recommender systems handbook. Springer.

[2] J.B.Schafer,D.Frankowski,J.Herlocker,S.Sen, Collaborative filtering recommender systems. The Adaptive Web, 291-324, 2007

[3] G. Adomavicius, A. Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749, 2005.

[4] M.J. Pazzani, A Framework for Collaborative, Content-based and Demographic Filtering. Artificial Intelligence Review, 13(5-6), 393-408, 1999

[5] Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. KnowledgeBased Systems, 46, 109-132.

[6] Han, J., Kamber, M., & Pei, J. (2011). Data mining: concepts and techniques: concepts and techniques. Elsevier

[7] M. O. Conner and J. Herlocker. Clustering Items for Collaborative Filtering. In Proceedings of the ACM SIGIR Workshop on Recommender Systems, Berkeley, CA, August 1999.

[8] S.H.S. Chee, J Han, K. Wang. Rectree: An efficient collaborative filtering method. Lecture Notes in Computer Science, 2114, 2001

[9] L. H. Ungar and D. P. Foster. Clustering Methods for Collaborative Filtering. In Proc. Workshop on Recommendation Systems at the 15th National Conf. on Artificial Intelligence. Menlo Park, CA: AAAI Press.1998

[10] Yuanyuan Wang, Stephen Chi-fai Chan and Grace Ngai Applicability of Demographic Recommender System to Tourist Attractions : A Case Study on Trip Advisor , 2012 IEEE DOI 10.1109/WI-IAT.2012.133

[11] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering, Proceedings of the Fifth International Conference on Computer and Information Technology, 2002

[12] A. Said, T. Plumbaum, W. E. De Luca, and S. Albayrak, “A comparison of how demographic data affects recommendation,” in Proc. 19th international conference on User modeling, adaption, and personalization, 2011.