Journal of Network Communications and Emerging Technologies (JNCET) www.jncet.org
Volume 5, Special Issue 2, December (2015)

A Novel Approach for Recommendation Based On Collaborative Filtering

Sangeeta
Department of Computer Engineering, YMCA University of Science & Technology Faridabad, India.

Dr. Neelam Duhan
Department of Computer Engineering, YMCA University of Science & Technology Faridabad, India.

Amrita Malhotra
Department of Computer Engineering, YMCAUST, Faridabad, India

Abstract – Recommendation is a technique used for determining an individual's interest in products, services, e-reading etc. Based on that interest recommendations are given to a target user. There are various approaches present in the literature for recommendation but generally collaborative filtering, content-based filtering and hybrid methods are used. All these techniques suffer from some drawbacks such as cold start, sparsity, accuracy etc. Collaborative filtering is the most widely used method for recommendation because of ease and quality. This paper provides a novel approach which solves some of the drawbacks in the technique and provide better recommendations.

Index Terms – collaborative filtering, recommender systems, similarity computation, item groups etc..

This paper is presented at International Conference on Recent Trends in Computer and information Technology Research on 25th& 26th September (2015) conducted by B. S. Anangpuria Instituteof Technology & Management, Village-Alampur, Ballabgarh-Sohna Road,Faridabad.

1. INTRODUCTION

Recommender systems find user's interest by various techniques and then suggest some items or services to the user. Recommender systems are widely used in e-commerce and social websites. For example YouTube, Facebook, amazon etc. Many e-retailers embedded recommendation systems in their web sites, in order to study the tastes of their customers, and achieve some business objectives. There are three techniques available for recommendation.

- Collaborative filtering
- Content based filtering
- Hybrid filtering

Collaborative filtering find interest using ratings given by users. Content based methods learns by some training data and then provide recommendations using the results of training dataset. Content based method uses information from user profiles such as profession, age, sex etc and use this information to find their interest [1]. Hybrid methods make use of both the approaches which overcomes most of the advantages of collaborative and content based filtering. In this paper collaborative filtering is explained in detail.

1.1 Collaborative Filtering

Collaborative filtering algorithm [2] is the most researches method in recommender systems. Collaborative method collects user's ratings as their feedback and use these ratings to predict new items or services they may like. Collaborative filtering works in various domains like restaurants, e-commerce, newspaper articles, movies etc. In this paper movie domain is considered to explain the proposed approach. Collaborative filtering finds users in a community that share appreciations [3]. If two users have same or most of same rated items in common, then they have similar tastes [4]. Such users build a group. Collaborative methods can be implemented in two ways.

- User based methods
- Item based methods

1.1.1.User based method

These algorithms consider similar user profiles to the target user and then items liked by these similar users are recommended to the target user. User based collaborative filtering is similar to the nearest neighbor method. This technique finds the nearest neighbor of target user. It finds

users with similar taste of target user. In this method, first step is to find nearest neighbor for which we need to obtain the users history profile. By analyzing the history, a rating matrix can be prepared in which each entry represents the rating of the user given to an item [5]. Each row in a matrix represents individual user and column represents an item, and the number at the intersection of a row and a column represents the user's rating value. If a user has not yet rated the item, intersection of that row and column is empty. The second step is to compute the similarity between target users and find their nearest neighbors. The Pearson correlation coefficient method is the most widely used for similarity computation.

User-based methods utilize entire database to find recommendations [6]. This approach is very popular in past but now there are other alternatives also which provide better recommendation. Main challenge of the approach is that user item matrix is very sparse and recommendations based on the matrix are of poor quality.

1.1.2. Item based methods

Item based method is a collaborative filtering technique which analyzes the set of items target user has rated and calculates how similar they are. After computing similarity of all rated items by target user, k most similar items are selected for prediction. Similarity and prediction computation is given below in detail.

a) Item Similarity

This is most crucial step in item based collaborative technique. Similarity computation largely effects quality of recommendation. Here are three methods for item similarity computation.

- Cosine based similarity
- Co-relation based similarity
- Adjusted cosine based similarity

i) Cosine based similarity:

In this method, two movies are thought of as different vectors. The similarity between these two vectors can be determined by calculating cosine angle between these vectors [7]. Let us assume two movies i and j as vectors then the similarity between i and j is given by (1).

$$\text{Sim(i,j)} = \cos(i,j) = \frac{i \times j}{\|i\| \|j\|} \qquad (1)$$

ii) Co-relation based similarity:

In this case similarity between two movies i and j is calculated by using Pearson-r Correlation [8]. Firstly isolate the co-rated cases, then compute the Pearson correlation based similarity as in (2).

$$\text{Sim(i,j)} = \frac{\sum_{a \in U_{ij}} (R_{a,i} - \overline{R_i})(R_{a,i} - \overline{R_j})}{\sqrt{\sum_{a \in U_{ij}} (R_{a,i} - \overline{R_i})^2} \sqrt{\sum_{a \in U_{ij}} (R_{a,i} - \overline{R_j})^2}} \qquad (2)$$

Where $R_{a,i}$ denotes the rating of movie i by user a. $\overline{R_i}$ and $\overline{R_j}$ are average rating of movies i and j.

iii) Adjusted cosine based similarity:

Adjusted cosine based similarity provides better quality similarity because it considers user's average rating for each co-rated pair [9]. This will overcome the difference between rating scales of individual user. Mathematically similarity is given as in (3).

$$\text{Sim(i,j)} = \frac{\sum_{a \in U_{ij}} (R_{a,i} - \overline{R_a})(R_{a,j} - \overline{R_a})}{\sqrt{\sum_{a \in U_i} (R_{a,i} - \overline{R_a})^2} \sqrt{\sum_{a \in U_i} (R_{a,j} - \overline{R_a})^2}} \qquad (3)$$

where $R_{a,i}$ and $R_{a,j}$ denotes the rating of movie i and j by user a. $\overline{R_a}$ are average rating given by user a.

b) Item Prediction

After computing similarity between two items prediction can be calculated to provide top N recommendation. Prediction can computed using weighted sum [10] method given in (4).

$$\text{P(u,i)} = \frac{\sum_{j \in NI} sim(i,j) * R_{u,j}}{\sum_{j \in NI} |sim(i,j)|} \qquad (4)$$

where sim(i,j) represents similarity between i and j calculated from (1). $R_{u,j}$ represents rating of item j given by user u. P(u,i) gives prediction of item i for target user u.

2. PROPOSED RECOMMENDER SYSTEM

The proposed system is based on collaborative filtering technique that is used to provide item recommendations to a target user by analyzing its interest. Collaborative filtering technique aggregates ratings on given items by a number of users and finds the user-user similarity based on these ratings which is further used to computes prediction weights. The detailed architecture of the purposed system and explanation of its functional components is given in the subsequent sections.

2.1. The Proposed System Architecture

The detailed system architecture is outlined in fig. 1.1 along with all functional components. In the proposed system, User interacts with an online interface by signing in or logging in. A new user need to register himself or can have a guest view on the website. An existing user can log in or use the guest

view. After log in or sign up process, online interface interacts with movie database which in turn stores movie rating matrix and movie genre matrix. Group generator generates groups of movies based on their genre information using movie database. After generating different movie groups, movie group rating generator generates group rating matrix using movie-user rating matrix and then stores the result in movie database. After generating group rating matrix user-user similarity between users interacts is computed by Pearson-correlation formula using group rating matrix, the results of which are stored in database. After calculating similarity information, traditional method for prediction is used to predict most suitable movies for a particular user using user-user similarity and movie-user matrix.

Predicted values are then sorted in decreasing order. Top N predicted values are then stored in movie repository where N=10. Movies information about predicted values is extracted from database and displayed to the target user.

2.2. Functional Components of Proposed System

The description of functional components of the proposed recommender system is given below.

a) Group generator

Group generator generates groups of movies with similar genres. It takes input from the movie repository and use movie-genre matrix to generate different groups. These groups are then used by group rating generator to produce group rating matrix.

b) Group rating generator

Group rating generator produces group rating matrix by reading user-movie matrix A from movie repository. It uses groups of movies generated by group generator and form a matrix which gives rating given by particular user to a group. If user rating of a particular group is higher as compared to other groups than the user's interest can be identified with in those groups. Group matrix is produced by averaging the ratings of similar movies given by a user. Values of matrix are the computed using the formula given in (5).

$$R_{u,C} = \frac{\sum_{i\in 1}^{C} R_{u,i}}{|R_{u,i}|} \qquad (5)$$

where $R_{u,C}$ is the rating of C^{th} cluster for user u. i represents movies in C^{th} cluster. $R_{u,i}$ denotes rating of i^{th} movie given by user u. $|R_{u,i}|$ denotes number of movies in C^{th} cluster rated by user u. The value of $R_{u,C}$ ranges in [1,5].

c) Similarity calculator

The amount of correlation between the users needed to calculate for similarity computation. Correlation between any two users is computed using Pearson-correlation formula [11,14] which gives efficient result. In statistics, the value of the correlation coefficient varies between +1 and -1. When the value of the correlation coefficient lies around ± 1, then it is said to be a perfect degree of association between the two variables[15]. Similarity calculator takes input from group rating matrix and user-movie matrix from movie repository and computes similarity between two users. This similarity signifies that how much two users have similar taste in movies. Similarity is computed as given in (6).

$$\text{Sim(u,v)} = \frac{\sum_{i\in I}(R_{u,i}-\overline{R_u})(R_{v,i}-\overline{R_v})}{\sqrt{\sum_{i\in I}(R_{u,i}-\overline{R_u})^2}\sqrt{\sum_{i\in I}(R_{v,i}-\overline{R_v})^2}} \qquad (6)$$

Where Sim(u,v) represents similarity between user u and v. $R_{u,i}$ denotes the rating of i^{th} group by user u. $R_{v,i}$ denotes the rating of i^{th} group given by user v. $\overline{R_u}$ and $\overline{R_v}$ are average rating of i^{th} group given by user u and v respectively . I is set of all users.

Similarity values ranges from 0 to 1 where 0 denotes user u and v does not share common interest. Value 1 denotes user u and v have same interest in movies.

d) Predictor

Predictor calculates weights of a movie for recommendation[13]. It uses traditional formula to compute prediction. Predicted values are stored in database. Predicted values ranges from 1 to 5. The prediction computation formula [11] is given as (7).

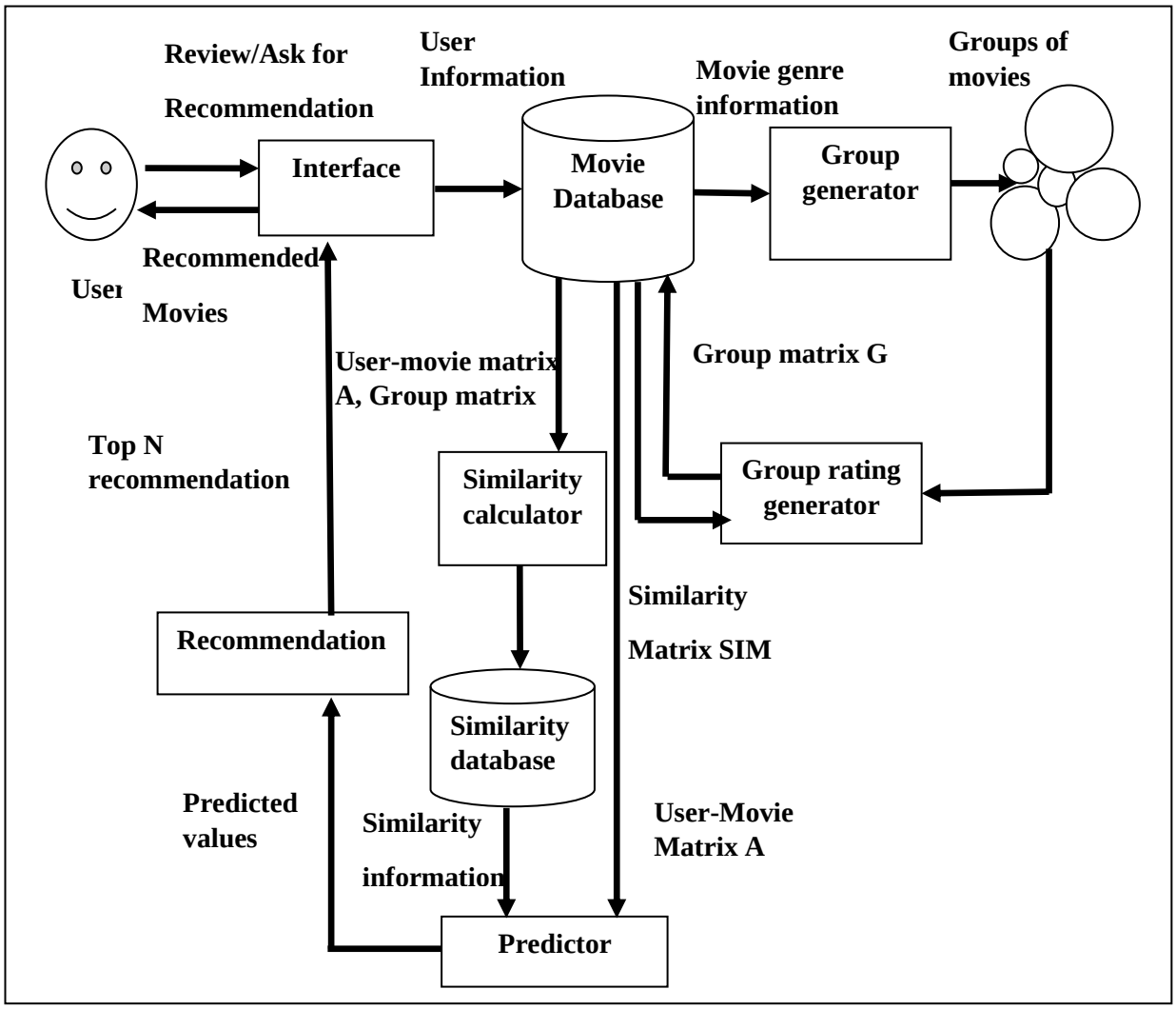
$$\text{P(u,i)}=\overline{R_u} + \frac{\sum_{v\in NI} sim(u,v)*R_{v,i}}{\sum_{v\in NI}|sim(u,v)|} \qquad (7)$$

where sim(u,v) represents similarity between user u and v calculated from (1.5). $R_{v,i}$ represents rating of i^{th} item given by user v, P(u,i) gives prediction of item i for target user u. value 5 represents the movie is strongly recommended. Value 1 denotes movie should not be recommended. Predicted values are then sorted in decreasing values and stored in an array list pre. NI is the set of movies which has not been rated by the target user.

e) Top n recommender

Predicted values are stored in a list in descending order. Now top 10 values is selected for recommendation to the target user. Value of N can be altered according to requirement. These top N values are then linked to movie repository and

corresponding movie information is retrieved and displays to the target user through online interface.



2.3. Accuracy of the System

Accuracy of the system is measured with the help of mean absolute error formula [12].Mean absolute error is a formula to find out the error in the system. Mean Absolute Error (MAE) between ratings and predictions is a widely used metric. MAE is a measure of the deviation of recommendations from their true user-specified values. Formally, it is given as (8).

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N} \quad (8)$$

where p_i is the predicted rating of i^{th} movie, q_i is actual rating of i^{th} movie and N is number of movies. The lower the MAE, the more accurately the recommendation engine predicts user ratings. The proposed system is checked for accuracy using this measure which shows that that the system is 97% accurate in terms of its predicted rating for user.

3. ISSUES SOLVED BY PROPOSED SYSTEM

Proposed system solves some of drawbacks in the existing system which are explained in detail in next section.

a. New item problem

When a new item is inserted in the system then it is added to the user-movie matrix A and movie-genre matrix B. It does not have any prior rating by the user so, it can not be recommended to the users. This is called new item problem.

Solution: When new movie is added to the system, it is checked with its type of genres and added to the groups of movies with similar genres. So, when a particular group of movie is recommended to the user, new item is also recommended with them.

b. New user problem

When a new user is inserted in the system it is added to user-movie matrix A. The user has not rated enough movies so that his/her interest can be calculated. So it is difficult to recommend such a user. This is known as new user problem.

Solution: A new user is added to the system in user-movie genre matrix A. for recommendation to such a user, top-n predictor calculates most famous movies rated by the user by computing average rating of a movie. These predictions are recommended to the new user until they rate enough movies to find their interest and personal recommendation are given after that.

c. Accuracy

Quality of recommended items is an important issue in recommender systems. Recommender system should provide maximum accuracy.

Solution: Proposed system is measured with the metric MAE to check the accuracy of system. Experimental results shows that system is more accurate than other techniques.

d. Sparsity

Collaborative filtering consider user-item matrix A for user/item similarity computation. This matrix is always sparse. Values rated by users are very less as compared to number of movies. So, it effects quality of recommendation.

Solution: Proposed system used Group Rating matrix for similarity computation which is not sparse and give efficient results.

4. CONCLUSION AND FUTURE WORK

This paper proposes an approach based on collaborative filtering which solves sparsity and accuracy problem. The proposed recommender system considers a movie database downloaded from movielens website. The proposed recommender system forms movie groups based on genre information. Whenever a new movie is entered in the system, it is checked with corresponding genre information and added to the relevant group. When recommendations are made from the group, the newly added movie is recommended with the other movies which are not rated by the target user and having the high prediction value. Whenever a new user is entered in the system, he/she is recommended with most famous movies rated by other user. Group rating matrix is formed by using group of movies and movie database. Recommendations are made by computing user to user similarity with the help of Pearson correlation method and then predictions are computed. These predictions are sorted in descending order and Top 10 recommendations are displayed to the user. This system solves various issues like sparsity, accuracy, cold start etc.

In Future Groups of movies can be improved using various features of movie such as director, release date, casting information etc. Similarity computation is done with the help of Pearson Correlation method which is available in literature. An improved correlation method can be developed to increase accuracy.

REFERENCES

[1] P. Melville, R. J. Mooney, and R. Nagarajan, “Content-boosted collaborative filtering for improved recommendations,” in Proceedings of the 18th National Conference on Artificial Intelligence (AAAI '02), pp. 187–192, Edmonton, Canada, 2002.

[2] Cover, T.,and Hart, P., Nearest neighbor pattern classification. Information Theory, IEEE Transactions on, 13(1):21–27, 1967.

[3] A. Elgohary, H. Nomir, I. Sabek, M. Samir, M. Badawy, and N. A.Yousri, “Wiki-rec: A semantic-based recommendation system usingwikipedia as an ontology,” in Intelligent Systems Design and Applications(ISDA), 2010 10th International Conference on, 29

[4] K. O. et al. Context-aware svm for context-dependent information recommendation. In International Conference On Mobile Data

[5] Zhi-Dan Zhao, Ming-Sheng Shang,”User-based Collaborative-Filtering Recommendation Algorithms on Hadoop”. 2010 Third International Conference on Knowledge Discovery and Data Mining.

[6] Xiangwei Mu, Yan Chen and Taoying Li,”User-Based Collaborative Filtering Based on Improved Similarity Algorithm”. 978-1-4244-5540-9/10/$26.00 ©2010 IEEE.

[7] Liang Hu, Guohang Song, Zhenzhen Xie, and Kuo Zhao,”Personalized Recommendation Algorithm Based on Preference Features”. TSINGHUA SCIENCE AND TECHNOLOGY ISSNll1007-0214ll08/11llpp293-299 Volume 19, Number 3, June 2014.

[8] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl,”Item-Based Collaborative Filtering Recommendation Algorithms”. WWW10, May 1-5, 2001, Hong Kong.ACM 1-58113-348-0/01/0005.

[9] Liang Hu, Guohang Song, Zhenzhen Xie, and Kuo Zhao,”Personalized Recommendation Algorithm Based on Preference Features”. TSINGHUA SCIENCE AND TECHNOLOGY ISSNll1007-0214ll08/11llpp293-299 Volume 19, Number 3, June 2014.

[10] Junhao WEN and Wei ZHOU,”An Improved Item-based Collaborative Filtering Algorithm Based on Clustering Method”. Journal of Computational Information Systems 8: 2 (2012) 571-578.

[11] Linden, G., B. Smith, and J. York. Amazon.com Recommendations,"Item-to-Item Collaborative Filtering". IEEE Internet Computing, Jan.-Feb. 2003.
[12] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: A constant time collaborative filtering algorithm". Information Retrieval, vol. 4, no. 2, pp. 133–151, July 2001.
[13] Francesco Ricci and Lior Rokach and Bracha Shapira, "Introduction to Recommender Systems Handbook, Recommender Systems Handbook". Springer, 2011, pp. 1-35.
[14] Jun Wang1 , Arjen P. de Vries , Marcel J.T. Reinders,"Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion". SIGIR'06, August 6–11, 2006, Seattle, Washington, USA Copyright 2006 ACM 1-59593-369.
[15] Luis M. de Campos, Juan M. Fernández-Luna and Juan F. Huete,"A collaborative recommender system based on probabilistic inference from fuzzy observations". Fuzzy Sets and Systems 159 (2008) 1554 – 1576.