International Journal of Applied Engineering & Technology

RECOMMENDATION SYSTEM USING HYBRID SIMILARITY MEASURE

Dr. Vimalkumar B. Vaghela
Assistant Professor, Department of Computer Engineering, L. D. College of Engineering, Ahmedabad, Gujarat, India
vimalvaghela@gmail.com

ABSTRACT
Collaborative filtering is one of the most successful and widely used methods of product recommendation to users of online store. The most critical component of this method is finding similarities among users using user-item rating matrix so that the product can be recommended to the user based on similarities. The varieties of measures used for finding similarities are Cosine, Pearson correlation coefficient, mean squared difference, etc. An important issue of recommendation system that is viewed by researchers is the new user cold start problem, which occurs when a new user is coming to the system and less rating is available in user-item rating matrix. The mentioned measures are not suitable for new user cold start problem. For the new user cold start problem, here we discus about the Hybrid similarity measure and also derive the user-user similarity matrix for the said measure. For an active user we find the top k neighbors and then according those neighbors we predict the missing ratings for an active user and then recommending a movie to a new active user.

Keywords: recommendation system, similarity measure, collaborative filtering, content based filtering, cold start

1. INTRODUCTION
Nowadays, more and more people have their own tablet and intelligent terminals and they are spending their time in accessing all types of social networking sites. Lots of information available about the products, but sometimes they are not happy with the result they get about the products. Moreover what the user is searching and what are his/her preferences are recorded on to the social media sites. Based on this context, the recommendation system is used to recommend the items to a user according to his/her preferences. Collaborative filtering is the most useful approach used by recommendation systems that provide personalized services to the users. An important issue for the RS that has captured the attention of researchers is the new user cold start problem. According to this problem, when the new user enters in the system, no prior ratings are available for this user in user-item rating matrix. And that is why the accuracy of recommendation system decreases. So these things motivate us to improve the accuracy of recommendation system under the cold start problem.

Nowadays, more and more people have their own tablets and other intelligent terminals. They are spending their time in accessing all types of social networking sites. Lots of information are available about the products, but sometimes they are not happy with the result they get about the products [2]. Moreover what the user is searching and what are his/her preferences are recorded on to the social media sites. Based on this context, the recommendation system is used to recommend the items to a user according to his/her preferences. Thus, The recommendation system use the information about users, user profiles to predict the utility or relevance of particular item. In such a way, they provide personalized recommendations. The recommendation systems have proven to be useful in the area such as e-commerce and they are having future in other domain like web search engine, digital TV program recommenders, etc [4].

2. RELATED WORK
The collaborative filtering has become the most widely used method to recommend items for users. It makes recommendation according to the similar users with the active user or the similar items with the items which are rated by the active user. It includes memory based/ user based and model based methods [6]. The content based filtering recommend the items based on their content where user profile is in the form content in which the user is interested. For example, the keywords of purchased books of a user are used to find other books that contain the similar keywords.

Content based filtering is unable to evaluate the quality of an item. It cannot distinguish the bad items from a good one [6]. While collaborative filtering is not so sensitive to the problems that are mentioned with content based filtering. The collaborative filtering will recommend items that have received high ratings by other users with same interests.

In collaborative filtering, the user profile is nothing but the set of ratings that are given to items by different users. These ratings are stored by observing the behavior of a user with the system [4].

The memory based method first calculates the similarities among users based on user-item rating matrix and then select the most similar users as the neighbors of an active user. These methods are using similarity measures to select users or items that are similar to an active user [2]. And last, it gives the recommendation to an active user according to the neighbors. These algorithms are also known as neighborhood based or user based algorithms. The memory based method gives good accuracy for recommendation, but the problem with this method is the computational time will grow if the size of user-item rating matrix increases. Some the following are with the memory based algorithms [3]:

Sparciry- In recommender systems, not all the users rates all the items from the available items. So many of the cells which are inuser-item rating matrix are empty. So in such a case, finding similarity among users or items becomes difficult [6].

Cold Start- In recommender system, when a new user enters in the system, no enough ratings are available for that user. So it is not possible to recommend the items to an active user [6].

Shilling- The recommender system is also suffering from the spam attacks due to the users that are interested in misleading the system to recommend a certain product [6].

3. SIMILARITY MEASURES

The memory based collaborative filtering method uses similarity measures to find the similarities among users or items. With the help of these measures we can calculate the similarities among users and some of the measures are as follows [4].

Sr. No.	Name of measure	Description
1	Pearson Correlation Coefficient	It is working on set of common rated items by both the users. Consider the absolute ratings.
2	Cosine measure	It is working on set of common rated items by both the users. Consider the absolute ratings.
3	Jaccard measure	It considers only number of common ratings between two users and not consider the absolute ratings
4	Proximity-Impact-Popularity measure	It denotes how common two user's ratings have.
5	New heuristic similarity measure	It punishes the small proportion of common ratings and considers the preferences of each user. It is normalized in (0,1)

4. A USER BASED SIMILARITY MODEL USING FUSION SIMILARITY MEASURE

4.1 Overview of Methodology

1 Input review dataset

2 Preprocessing Technique

Creation of user-movie rating matrix

3 Applying Hybrid Similarity measure

4 Generation of user-user similarity matrix

5 Find top k nearest neighbor

6 Apply prediction on ratings

7 Recommend movies to an active user

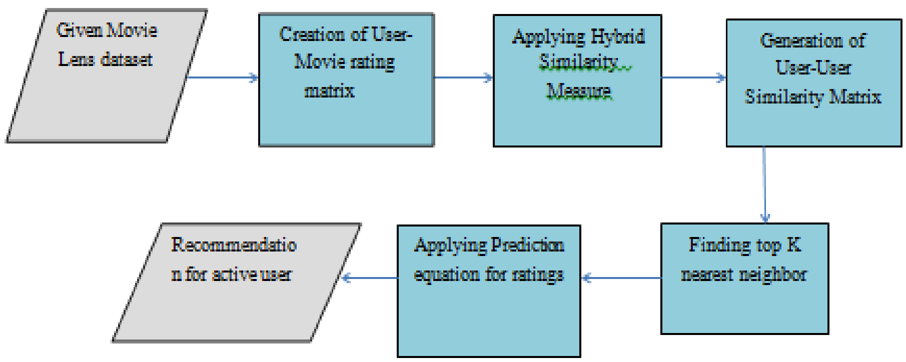


Fig.1 Flow of Proposed System Architecture

4.2 Algorithm

Algorithm: RecSys (UId, MovId, Rating)

Input:
Uid: User id of movie viewer who has rated the movies.

MovId: Unique Identification of movie.

Rating: Rating in the scale of [1-5] given by viewer.

Output:
Rated movie list according to user preference.

Procedure:
Step 0: [pre-processing]

Convert xls-database to array User-MovieArray

Step 1:
SimilarityMatrix=User-User_SimilarityMatrixGen(User-MovieArray)

Step 2:
KNNProc(UId, MovId)

Step 3:
Call Rating_Predict(UId, MovId)

Step 4:
Recommend rated MovieList

Step 5:
End

Proc User-User_SimilarityMatrixGen (User-MovieArray)

{

For i=1 to User-MovieArray.rowlength

For j=i+1 to User-MovieArray.rowlength

Compute the similarity between User u=u_i and User v=$u_j$for common

rated items

Compute $sim(u,v)^{NHSM} = sim(u,v)^{JPSS} . sim(u,v)^{URP}$

Where:

$$sim(u,v)^{URP} = 1 - \frac{1}{1+\exp(-|\mu_u - \mu_v| . |\sigma_u - \sigma_v|)}$$

$$sim(u,v)^{JPSS} = sim(u,v)^{PSS} . sim(u,v)^{Jaccard}$$

Where:

$$sim(u,v)^{Jaccard} = \frac{|I_u \cap I_v|}{|I_u| \times |I_v|}$$

$$sim(u,v)^{pss} = \sum_{p \in I} PSS(r_{u,p}, r_{v,p})$$

End

End

Return SimilarityMatrix()

}

Proc KNNProc(UId,MovId)

{

Find the top 5 nearest neighbors for all movie id's in a User-Movie Rating

Matrix for a given UId by using the Similarity Matrix find in above step.

Nearest neighbors are those users who are having the maximum similarity with

a given User Id (UId) .

}

Proc Rating_Predict(UId, MovId)

{

Make prediction of ratings for a given UId for all movie id's in a User-Movie

Rating Matrix with the help of the nearest neighbors of an active user UId.

For i=1 to User-MovieArray.columnlength

Compute $P_{a,i} = \frac{\bar{r}_a + \sum_{u=1}^{n} (r_{u,i} - \bar{r}_u) * w_{a,u}}{\sum_{u=1}^{n} w_{a,u}}$

Where:

$$w_{a,u} = \frac{\sum_{i=1}^{m}(r_{a,i} - \acute{r}_a)(r_{u,i} - \acute{r}_u)}{\sigma_a \sigma_u}$$

End

}

5. RESULTS and discussions

For the implementation purpose, we used the Movie Lens dataset which contains the following information.

No of Records : 100000

No of Users : 943

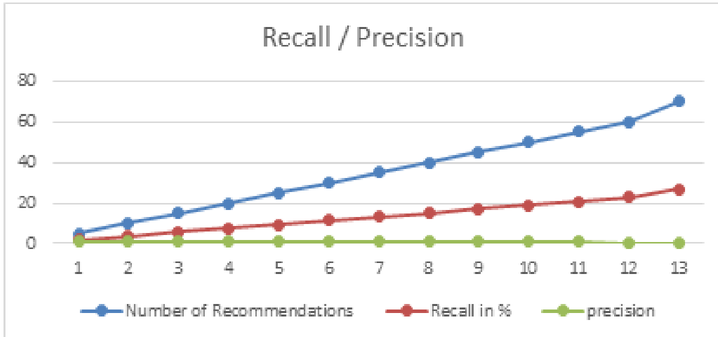
No of Movies : 1682

From this data, I extract cold start data. And according to that we fetch data for 1 to 20 user id and 1 to 50 movie id and perform the implementation on about 110 records.

$$\begin{array}{c} \\ u1 \\ u2 \\ u3 \\ u4 \\ u5 \end{array} \begin{array}{c} \begin{array}{ccccc} u1 & u2 & u3 & u4 & u5 \end{array} \\ \begin{pmatrix} 0.05992 & 0.02377 & 0.05033 & 0.00561 & 0.02034 \\ 0.02377 & 0.05856 & 0.02081 & 0.00593 & 0.03113 \\ 0.05033 & 0.02081 & 0.05992 & 0.00751 & 0.02084 \\ 0.00561 & 0.00593 & 0.00751 & 0.03412 & 0.01026 \\ 0.02034 & 0.03113 & 0.02084 & 0.01026 & 0.05672 \end{pmatrix} \end{array}$$

- From this similarity matrix, we can see that the similarity between user1 and user3 is higher than the similarity between user1 and user2 and it is according to user-item rating matrix given in Table 11. While for old similarity measures this is not true.
- The similarity between user3 and user5 is higher than the similarity between user4 and user5. While for PCC, COS, ACOS, Jaccard and PIP similarity this is not true.
- The similarity between user1 and user2 is higher than the similarity between user3 and user4. While for PCC, COS, ACOS and Jaccard similarity this is not true.
- From the similarity matrix, we can also say that each user has different similarity, so they are comparable with one another and this is not true for other measures.

QUALITATIVE PERFORMANCE OF RECOMMENDATIONS



6. CONCLUSION

In this study, we discuss about the new hybrid similarity measure. There is a discussion on cold start problem when a new user comes in the system, small numbers of ratings are available for that user and the recommendation accuracy is degraded. By taking this new hybrid measure, the user-user similarity matrix that we obtained is more accurate in comparison with other old similarity measures. So we can say that if we do the predictions and/or recommendations to an active user by using this hybrid similarity matrix, the recommendation accuracy is also increased in comparison with old existing similarity measures.

7. REFERENCES

[1] Han, Jiawei, and Micheline Kamber, Data Mining, Southeast Asia Edition: Concepts and Techniques. Morgan kaufmann, 2006.

[2] H.J. Ahn, A new similarity measure for collaborative filteringto alleviate the new user cold-starting problem, Inform. Sciences 178 (2008) 37-51.

[3] HYUNG JUN AHN , A Hybrid Collaborative Filtering Recommender System using a new similarity measure, Hangzhou, China , April 15-17 , 2007

[4] Haifeng Liu, Zheng Hu, Ahmad Mian, Hui Tian, uzhen Zhu : A new user similarity model to improve the accuracy of collaborative filtering ,Knowledge-Based Systems 56 (2014) 156–166

[5] Le Hoang Son : Dealing with the new user cold-start problem in recommender systems: A comparative review , Information Systems (2014)

[6] Carneiro, D. Fernández, V. Formoso, Comparison of collaborative filtering algorithms: limitations of current techniques and proposals for scalable, high- performance recommender system, ACM Trans. Web 5 (1) (2011) 1–33.

[7] Bidyut Kr. Patra, Raimo Launonen, Ville Ollikainen Sukumar Nandi : A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data Knowl. Based Syst. (2015)

[8] Ruchika, Ajay Vikram Singh, Dolly Sharma: Evaluation Criteria for measuring the performance of Recommender Systems, IEEE conference 2015

[9] Jonathan L. Herlocker, Joseph A. Konstan, Al Borchers and John Riedl: An Algorithmic Framework for Performing Collaborative Filtering, ACM Transactions

[10] J. Bobadilla, F. Ortega, A. Hernando, A. Gutierrez: Recommender systems survey, Knowledge-Based Systems 46 (2013) 109-132

[11] Hao Ma, Irwin King, Michael R.Lyu: Effective Missing data prediction for Collaborative filtering, SIGIR 2007 Proceedings

[12] Javad Basiri, Azadeh Shakery, Behzad Moshiri, Morteza Zi Hayat: Alleviating the Cold-Start Problem of Recommender Systems using a New Hybrid approach, 2010 5th International Symposium on Telecommunications

[13] Ke Zhou, Shuang-Hong Yang, Hongyuan Zha: Functional Matrix Factorizations for Cold start Recommendation, SIGIR'11, July 24-28, 2011, Beijing, China

[14] L. Safoury, A. Salah, Exploiting user demographic attributes for solving cold-start problem in recommender system, Lect. Notes Softw. Eng. 1 (3) (2013) 303–307

[15] P. Resnick, H.R. Varian, Recommender systems, Commun. ACM 40 (3) (1997) 56–58 B.N. Miller, I. Albert, S.K. Lam, J.A. Konstan, J. Riedl, MovieLens unplugged: experiences with an occasionally connected recommender system, in: proceedings of the 8th International Conference on Intelligent user interfaces, 2003, pp. 263–266

[16] J. Bobadilla, F. Ortega, A. Hernando, A collaborative filtering similarity measure based on singularities, Inform. Process, Manage. 48 (2012) 204–217

[17] J. Bobadilla, F. Ortega, A. Hernando, J. Bernal, A: collaborative filtering approach to mitigate the new user cold start problem, Knowledge-Based Syst. 26 (2011) 225–238

[18] J. Bobadilla, A. Hernando, F. Orteqa, A. Gutirrez, Collaborative filtering based on significances, Inform. Sci. 185(2012)1–17

[19] Mohsen Jamali, Martin Ester: TrustWalker-A Random Walk Model for Combining Trust-based and Item-based Recommendation, KDD'09, June 28–July 1, 2009

[20] Jadav, Bhumika M., and Vimalkumar B. Vaghela. "Sentiment analysis using support vector machine based on feature selection and semantic analysis." International Journal of Computer Applications 146.13 (2016).

[21] Vaghela, Vimalkumar B., Bhumika M. Jadav, and M. E. Scholar. "Analysis of various sentiment classification techniques." International journal of Computer applications 140.3 (2016): 975-8887.

[22] Vimalkumar B. Vaghela, "Feature Selection using Functional Dependency", Tuijin Jishu/Journal of Propulsion Technology 44 (4): 8019-8028

[23] Gandhi, Riya A., and VimalKumar B. Vaghela. "Novel approach to Case Based Reasoning System by aggregating Semantic Similarity Measures using Fuzzy Aggregation for Case Retrieval." *International Journal of Computer Applications* 163.10 (2017).

First A. Author Dr. Vimalkumar B. Vaghela, is an Assistant Professor in Department of Computer Engineering at L. D. College of Engineering, Ahmedabad, Gujarat, India. He had done his Ph.D. in Relational Data Mining. He has published more than 35 research papers in international journals and conferences. He had published a book titled "Ensemble Classifier in Data Mining". He has more than 18 years of teaching experience; his biography has been published in the Who's Who in Science and Engineering – 11th Edition, 2010-11, 2013-14, 2016-17. Written a Book with title "Operating System" in Dreamtech Publisher. He is Senior Professional Members of Association for Computing Machinery (ACM), Life time Professional Member of Computer Society of India (CSI) & Professional Member of International Association of Engineers (IAENG).