Advances in Computer Science and Information Technology (ACSIT)
p-ISSN: 2393-9907; e-ISSN: 2393-9915; Volume 3, Issue 4; April-June, 2016, pp. 310-313

http://www.krishisanskriti.org/Publication.html

Collaborative Filtering Recommendation System Based Upon User Reviews

Monika[1], Sanjeev Dhawan[2] and Kulvinder Singh[3]

[1]M.Tech. (Computer Engineering), Department of Computer Science & Engineering, University Institute of Engineering and Technology (U.I.E.T), Kurukshetra University, Kurukshetra, Haryana
[2]Faculty of Computer Science and Engineering , Department of Computer Science & Engineering, University Institute of Engineering and Technology (U.I.E.T), Kurukshetra University, Kurukshetra, Haryana
[3]Faculty of Computer Science and Engineering , Department of Computer Science & Engineering, University Institute of Engineering and Technology (U.I.E.T), Kurukshetra University, Kurukshetra, Haryana
E-mail: [1]monikachoudharylather@gmail.com, [2]kshanda@rediffmail.com, [3]rsdhawan@rediffmail.com

Abstract—*Nowadays, there are many recommendation systems, accessible via internet, which attempt to recommend to users several products such as music, movies, books, etc. Aiming at long response time and solving cold start problems that are faced by present recommendation algorithm. This paper, proposes a collaborative filtering approach based on user's credibility taking Movies as an example. This approach will find out the cluster that target user belongs to and further provide recommendation. Collaborative model will improve the response time, increased the performance and find out the Mean Absolute Error. Section first describes Introduction about Collaborative recommendation system, its work flow and why it is used. In second section related work about Collaborative Filtering. Section third describes how to find out the Mean Absolute Error and how to reduce it by using the Euclidean distance and Pearson correlation. And at last in forth section its experimental evaluations to predict the MAE and time to build the recommendation for each user.*

Keywords*: Collaborative Filtering, Pearson Correlation, Euclidean Distance.*

1. INTRODUCTION

Collaborative Filtering (CF) based approach help Users to make selection depend upon the opinions of other Users who share similar taste [1]. CF technique has further divided into two techniques, user-based and item-based CF technique [2]. In the user-based CF approach, a user will accept recommendations of movies (items) that are liked by users of identical interest. In the item-based CF approach, a user will accept recommendations of movies (items) that are similar to those movies or items which they have loved/liked in their past.

The workflow of a collaborative filtering system is described as below:

Firstly, a user provides rating to items (e.g. books, movies or CDs) by expressing his or her preferences. These ratings can be viewed as an approximate representation of the user's taste in the corresponding domain.

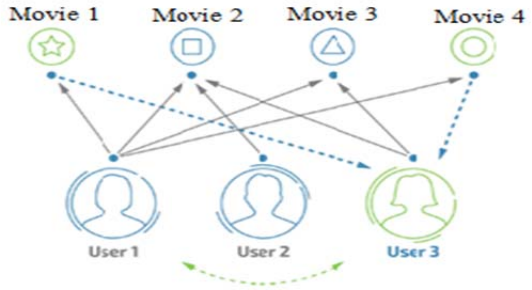

Fig. 1: General Process of Collaborative filtering

Secondly, the system matches the users' ratings against other users and finds the user who has "similar" tastes.

And at last, with similar users, the system recommends those items which have been rated highly by similar users but not yet being rated by active user (presumably the absence of rating is often considered as the unfamiliarity of an item).

Recommendation of movies is emerging with force nowadays due to the huge amount of movie content and because users normally don't have the time to search through these collections looking for new items.

The main purpose of a recommendation system is to estimate the user's preferences and present him with some items that he doesn't know yet.

It divided into two parts.

1.2 Memory Based Collaborative Filtering Technique

Memory-based collaborative filtering utilizes the user-item database in order to generate a prediction. On other hand, it finds the users who rated similar items or purchased similar

sets and by employing algorithms to combine the preferences of the neighbors; it produces the recommendation for the active user. Example: Nearest-neighbor algorithm.

1.2 Model-based Collaborative Filtering Technique

This technique depends on learning concept, that is, the system that can analyze the training data, summarize the complicated patterns into the learned models, and then make predications based on the learned models. The model building process can be processed by different machine learning algorithms such as Bayesian network, clustering and rule-based approaches.

2. RELATED WORK

There is an incessant growth in the Information Technology and web services on the internet. This growth of the internet has made it more difficult to the user to effectively extract information from all available online sources ,user have to spend a lot of time on the corresponding sources to extract useful information. The Grundy System presented by E.Rich [3] it was an initial step regarding automatic recommender systems which build model of individual user based upon very small information. Later on Tapestry system was discovered by Goldberg [4], which allowed the people to query for items in an information domain, such as corporate e-mail, based on other users' viewpoint or actions. Marlin [5] defined that Collaborative Filtering was work as a framework for filtering data established on the preferences of users.

Bell M. *and* Yehuda Koren [6] they have Concluded that cooperative filtering is an area established ("k-nearest neighbors"), whereas a user-item preference locale is interpolated from ratings of comparable items and/or users in past. In the same year, Robert M.Bell and Yehuda Koren[7]discovered neighborhood-based collaborative filtering in which they display how to derive simultaneously interpolation weights for all nearest acquaintances, unlike preceding ways whereas a single heaviness was computed separately.

Ahn and Hyung Jun [8] proposed a new methodology to reduce the user cold-starting problem. This can be done by calculating the similarities that depend on established distance and vector similarity measures such as Pearson's correlation and cosine which has been questioned about their effectiveness in last year's.

Heung-Nam Kim *et al.* [9] discovered that the proposed algorithm enhancing the recommendation quality for sparse data and in dealing alongside cold-start users as contrasted to continuing work. They analyzed the possible of cooperative tagging arrangements, encompassing personalized and biased user preference scrutiny, and specific and vibrant association of content for requesting the recommendations.

It is clear that although there are various techniques are available to improve the recommendations in order to attract a number of users. However, in most of the recommendation systems (like collaborative filtering, Content-Based and Hybrid), it is difficult to maintain the memory when network size becomes large. In order to avoid this problem, a methodology will be made to design the model-based collaborative filtering to define the recommendation based on User's Credibility.

3. RATING BASED ON SIMILARITY

3.1 Correlation-Based Similarity

In this case, similarity $w_{u,v}$ between two users u and v, or $w_{i,j}$ between two movies i and j, is measured by computing the Pearson correlation or other correlation-based similarities.

Pearson correlation measures the extent to which two variables linearly relate with each other [10]. For the user-based algorithm, the Pearson correlation between users' u and v is

$$w_{u,v} = \frac{\sum_{i \in I}(r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I}(r_{u,i} - \bar{r}_u)^2}\sqrt{\sum_{i \in I}(r_{v,i} - \bar{r}_v)^2}}$$

Where I ∈ I summations are over the items that both the users u and v have rated and $\bar{r}_u$u is the average rating of the co-rated items of the u_{th} user

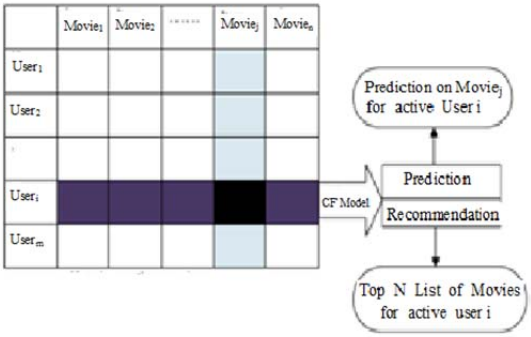

Fig. 2: User-based similarity ($w_{i,j}$) calculation based on the correlated items i and j from users 2, l and n.

For the item-based algorithm, denote the set of users' u ∈U who rated both items i and j, then the Pearson Correlation will be

$$w_{u,v} = \frac{\sum_{u \in U}(r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{i \in I}(r_{u,i} - \bar{r}_i)^2}\sqrt{\sum_{i \in I}(r_{v,j} - \bar{r}_j)^2}}$$

Where $r_{u,i}$ is the rating of user u on item i, $\bar{r}_i$ is the average rating of the ith item by those users, see Fig. 2 and 3.

Some variations of item-based and user-based Pearson correlations can be found. The Pearson correlation based CF

algorithm is a representative CF algorithm, and is widely used in the CF research community.

3.2 Vector Cosine-Based Similarity

The similarity between two Movies or Users can be calculated by treating each Movies or Users as a vector of word frequencies and computing the cosine of the angle

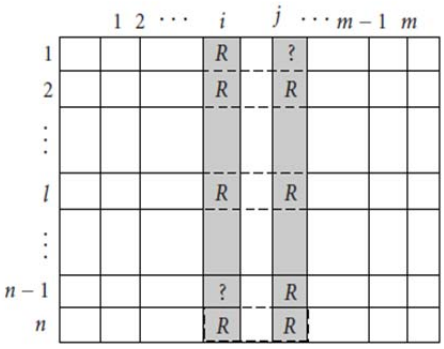

Fig. 3: Item-based similarity ($w_{i,j}$) calculation based on the correlated items i and j from users 2, l and n.

formed by the frequency vectors. This formalism can be adopted in collaborative filtering, which uses users or items instead of documents and ratings instead of word frequencies.

Formally, if R is the m × n user-movie matrix, then the similarity between two movies, i and j, is defined as the cosine of the n dimensional vectors corresponding to the i_{th} and j_{th} column of matrix R. Fig. 2 shows User-movie matrix.

Vector cosine similarity between movies i and j is given by

$$w_{i,j} = Cos(\vec{i},\vec{j}) = \frac{\vec{i},\bullet\vec{j}}{\|\vec{i}\| * \|\vec{j}\|}$$

where "•" denotes the dot-product of the two vectors. To get the desired similarity computation, for n movies, an n × n similarity matrix is computed. For example, if the vector $\vec{A}$= {x_1, y_1}, vector $\vec{B}$ = {x_2, y_2}, the vector cosine similarity between A and B is

$$w_{A,B} = Cos(\vec{A},\vec{B}) = \frac{\vec{A},\bullet\vec{B}}{\|\vec{A}\| * \|\vec{B}\|} = \frac{x_1x_2 + y_1y_2}{\sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}}$$

3.3 Euclidean Distance

Euclidean Distance is same as a simple distance between two points or users. Here we consider distance between two users i.e. length of path connecting them.

$$\text{E.D} = \sqrt{\sum_{i \in Ito\, n}(pi - qi)^2}$$

As an experiment we also show what will be Pearson correlation by taking both users based and item based example. And how many movies are reviewed by user and rating regarding to them that is given by users.

Closest user is calculated by Euclidean Distance and also specifies the smallest distance between all other users also.

3.4 Mean Absolute Error (MAE)

Instead of classification accuracy or classification error, the most widely used metric in CF research literature is Mean Absolute Error (MAE), which computes the average of the absolute difference between the predictions and true ratings.

$$MAE = \frac{\sum_{\{i,j\}}|p_{i,j} - r_{i,j}|}{n}$$

Where n is the total number of ratings over all users, pi, j is the predicted rating for user i on movie j, and $r_{i,j}$ is the actual rating. Lower the MAE, the better the prediction.

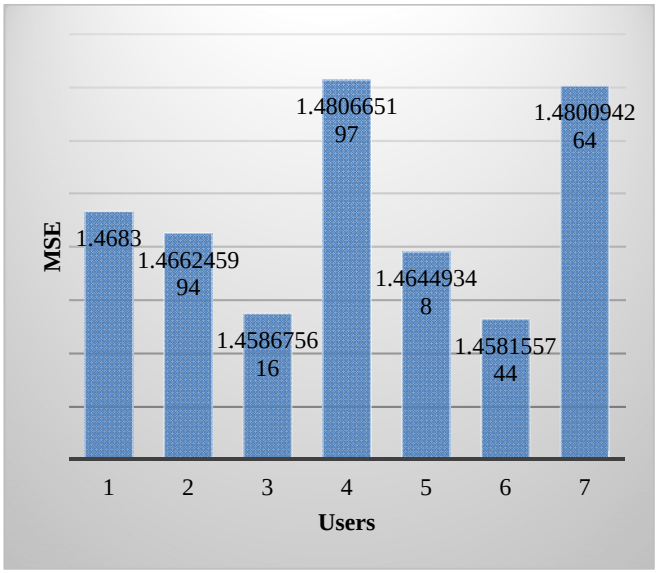

Fig. 4: User Specific MSE Using Euclidian Clustering for Movies

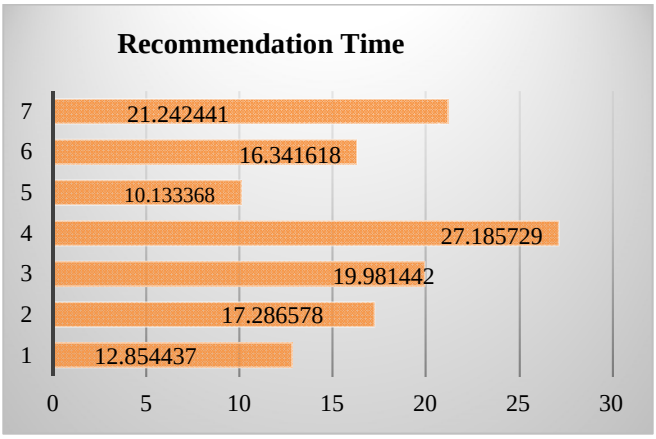

Fig. 5: Recommendation Time per user

4. CONCLUSION AND FUTURE SCOPE

Collaborative filtering aims at helping users to find movies that they should liked from big data. In that field, we can differentiate between various approaches like user-based and item-based. For each of them, many alternatives are available which are consider important to find their performances, for user- or item-based approaches similarity between users or items has been calculated and number of neighbors, the number of clusters for model-based approaches using clustering.

We have implemented collaborative filtering methodology based on user credibility and used real dataset called MovieLens to compare it, and using the same widely used performance measure called Mean Absolute Error(MAE).

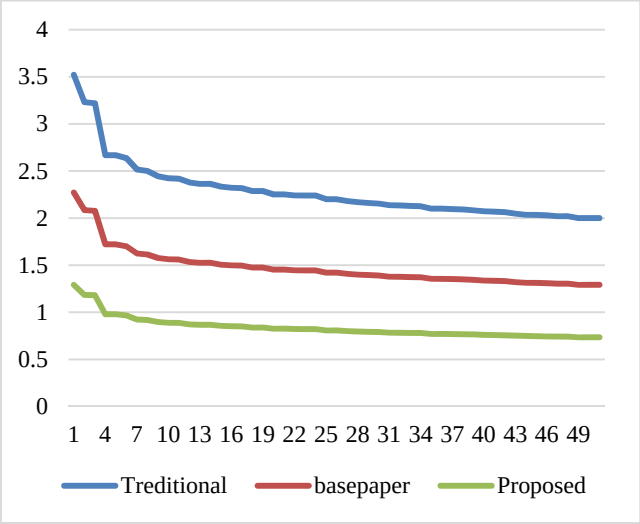

Fig. 6: Comparison of Response time using Traditional, Base Paper and Proposed work.

This work has thus allowed us to highlight the drawbacks of existing approaches and designed a new one. In this work, we have proposed new clustering based method for collaborative filtering, the proposed algorithm outperforms existing works by factor of 60%, and response time is enhance upto $3x_s$.

While a Mean Absolute Error of 1.415 presents improvement over result, it is still not good enough. However, since the two methods capture different sorts of information, we can get a better result by combining them. For example by taking a weighted average of the results (60% movie, 40% user), the RMSE can be lowered to 0.9030. More sophisticated combinations might go beyond using a fixed ratio, to modify the ratio across queries by estimating the relative success of the user-oriented method. In our future work, we will 1) investigate how to statistically quantify the "relatedness" between rating matrices in different domains, and 2) consider an asymmetric problem setting.

REFERENCES

[1]Koren Yehuda and Robert Bell, "Advances in Collaborative Filtering - Recommender Systems Handbook", pp. 145-185. Springer US, 2015.

[2]Badrul Sarwar, George Karypis, Joseph Konstan and John Riedl, "Item-based collaborative filtering recommendation algorithms" In the Proceedings of 10th international conference on World Wide Web, pp. 285-295, ACM, 2001.

[3]E. Rich, "User Modeling via Stereotypes", Cognitive Science, Vol. 3(4), pp. 329-354, 1979.

[4]D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," Communications of the ACM, vol. 35(12), pp. 61–70, 1992.

[5]Marlin Benjamin, "Collaborative filtering: A machine learning perspective" PhD diss., University of Toronto, 2004.

[6]Bell Robert M. and Yehuda Koren, "Scalable collaborative filtering with jointly derived neighborhood interpolation weights." In Data PMining, Seventh IEEE International Conference on, pp. 43-52. IEEE, 2007.

[7]Bell Robert M. and Yehuda Koren, "Improved neighborhood-based collaborative filtering", In KDD cup and workshop at the 13th ACM SIGKDD international conference on knowledge discovery and data mining, pp. 7-14, 2007.

[8]Ahn and Hyung Jun, "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem." Information Sciences 178, no. 1 , pp. 37-51, 2008.

[9] Kim Heung-Nam, Ae-Ttie Ji, Inay Ha, and Geun-Sik Jo, "Collaborative filtering based on collaborative tagging for enhancing the quality of recommendation" Electronic Commerce Research and Applications 9, no. 1 , pp.73-83, 2010.

[10]Das, Abhinandan S., MayurDatar, Ashutosh Garg, and ShyamRajaram. "Google news personalization: scalable online collaborative filtering." In*Proceedings of the 16th international conference on World Wide Web*, pp. 271-280. ACM,2007.