



Measuring Similarity Index for Collaborative Filtering Recommendation System

¹Monika, ²Kulvinder Singh, ³Sanjeev Dhawan

¹Student of M.Tech. (Computer Engineering), ^{2,3}Faculty of Computer Science & Engineering,

^{1,2,3}Department of Computer Science & Engineering,

University Institute of Engineering and Technology (U.I.E.T),

Kurukshetra University, Kurukshetra (K.U.K)-136119, Haryana, INDIA

Abstract: With the advancement in technologies, huge amount of information is present on the internet. Therefore to abstract relevant information from all information of user's interest a system is needed which is called as recommendation system (which recommend you the items, books, songs, movies and friends). Various recommendation systems are normally used in online social websites (like Twitter, Facebook, Instagram etc.) and also in variety of some most popular online shopping websites (like Amazon, e-bay, Flipkart, Myntra, Jabong, Snapdeal etc.). To perform recommendation a website know about the user profile (the user who want to buy the product) and the description of the highly rated item purchased or viewed by other user so that they can make recommendation fastly and accurately and the way that define that item is of user interest or not is depend on recommendation system. This will increase the rating of item which is highly recommend and increased the revenue of a company. Here collaborative recommendation technique is discussed to remove cold start problem and to minimize Mean absolute Error (MAE).

Keywords: Collaborative Filtering, Euclidean distance, Mean Absolute Error, Pearson correlation.

I. Introduction

Recommender systems have changed the view that how people will find products, information, and even other people. They study patterns of behavior to know what someone will prefer from among a collection of things he has never experienced. It will study the most important of those tools, including how they work, how to use them, how to evaluate them, and their advantages and drawbacks in practice.

1.1 Collaborative Filtering

Collaborative Filtering (CF) based recommendation techniques help people to make choices based on the opinions of other people who share similar interests Koren Yehuda and Robert Bell [1]. CF technique has divided into user-based and item-based CF approaches Badrul *et.al* [2].

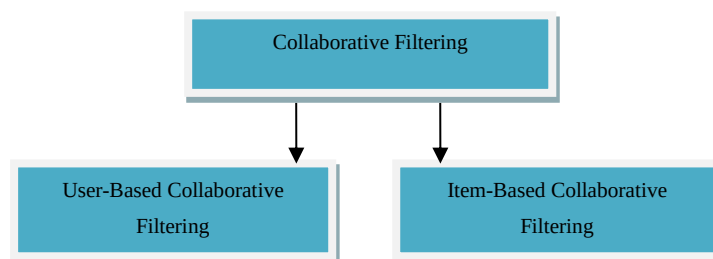


Figure 1: Types of Collaborative Filtering Recommendation System

II. Related Work

There is an incessant growth in the Information Technology and web services on the internet. To Find the relevant data from the online websites Recommendation System is used. Recommendation System provide user the items that are of their interest. The Grundy System presented by E.Rich [3] 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.

III. Problem Formulation

When Internet surfers search for information, they rely on recommendations from other people, customer reviews, or recommender systems. Recently, various kinds of recommender systems have attempted to reduce information overload and retain customers by providing personalized recommendations based on preferences.

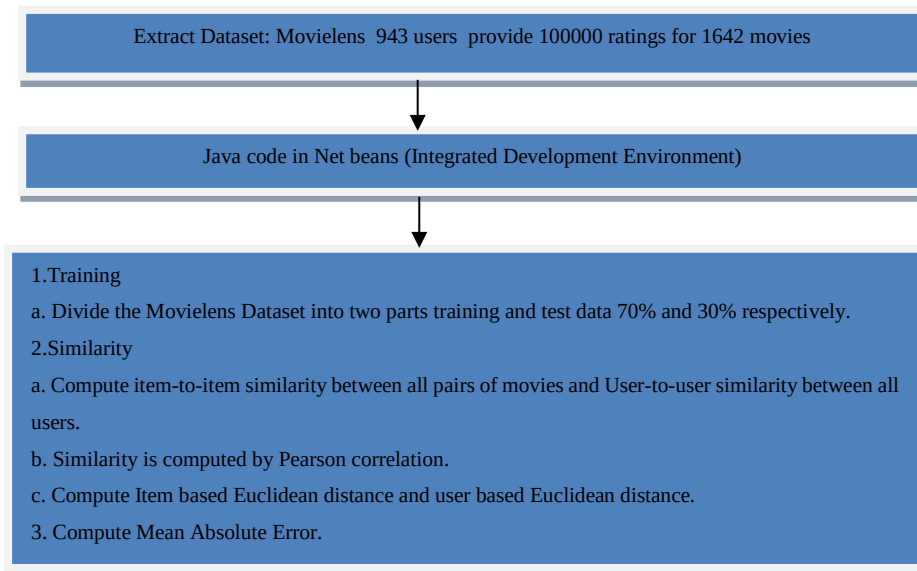


Figure 2 Workflow of Proposed work

These recommender systems use diversified algorithms to filter data and generate recommendations about items such as books, news, music, Web pages, and even virtual items. Among those various recommender systems, collaborative filtering (CF) has become the most popular recommendation algorithm. This system predicts user preferences based on the opinions of other similar users who have rated the items according to preference. By calculating the level of similarity between users in a rating data set, it becomes possible to find the nearest neighbors with the highest similarity values among all users. Once a user's neighborhood is identified, particular items can be evaluated by forming a weighted summation of neighbors' opinions.

IV. Experimental Evaluation

In order to evaluate the experiment done on Movielens Dataset which is available online at www.grouplens.org Laptop having Windows 7, 32 bit Operating System, Intel Core i3 Processor with 2GB RAM and Harddisk 460 GB is used. And the NetBeans Platform used For Coding need free disk space of 1.5GB only. Firstly Mean Absolute Error of user1, user2 and user3 is compute by Euclidean Distance. And then Pearson Correlation is used to find out the similarity between the users that helps to compute the MAE. Best Movie for user 87 will be find out by User based Pearson and Item based Pearson. Figure 3 depicts that MAE is find out by measuring the similarity between users by Euclidean Distance and the whole process is User based. Figure 4 state the MAE that is Item based is calculate between User1, User2 and User3 by measuring the similarity by Pearson Correlation. Figure 5 demonstrate that Pearson Correlation Similarity Measuring Technique is used for Item based Method to find the MAE. Figure 6 shows which Movie is best for User 87 by different methodology (User based and Item based).

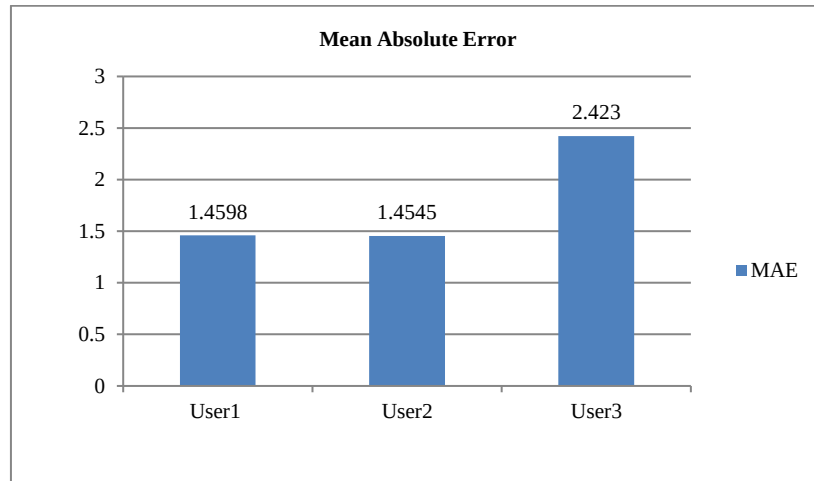


Figure 3 Mean Absolute Error (User based Euclidean Distance)

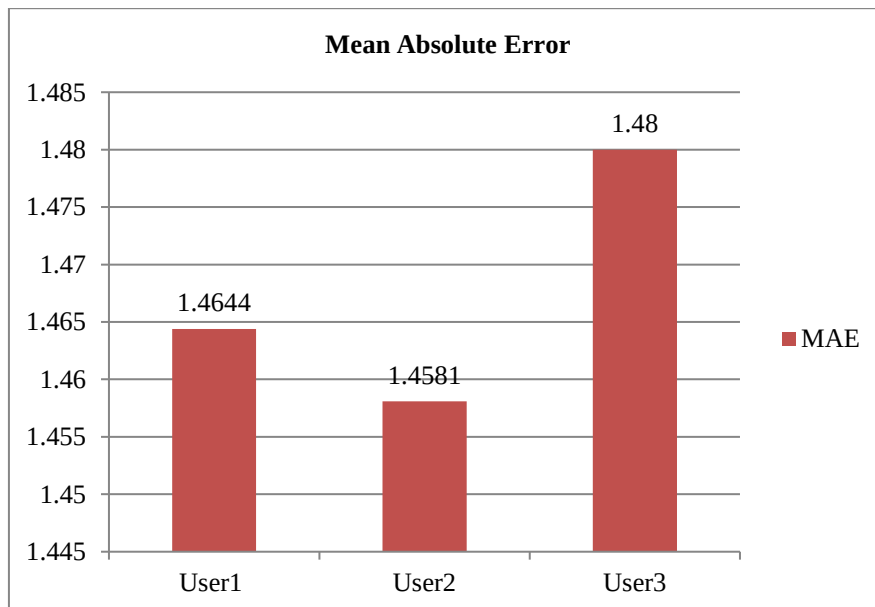


Figure 4 Mean Absolute Error (User-based Pearson)

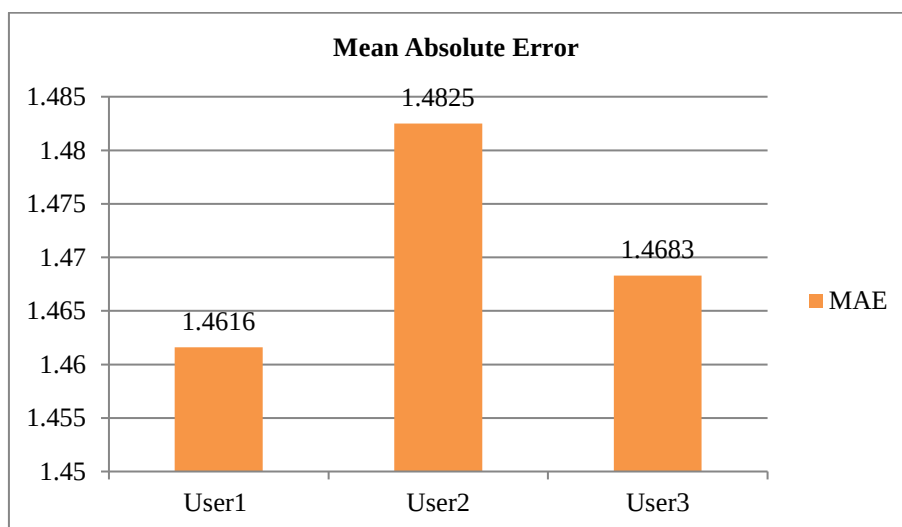


Figure 5 Mean Absolute Error (Item-based Euclidean)

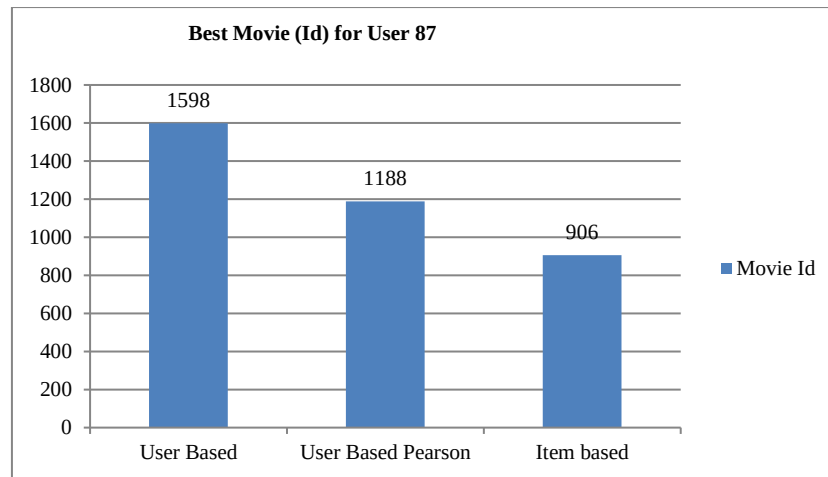


Figure 6 Best Movie for User 87

V. Conclusion

Collaborative filtering aims at helping users to find movies that they should liked from big data. In that field, differentiation 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. An improved collaborative filtering methodology based on user credibility is proposed which uses real dataset called MovieLens to compare it, and using the same widely used performance measure called Mean Absolute Error (MAE). In future work, an attempt will be made to investigate how to statistically quantify the “relatedness” between rating matrices in different domains, and also to 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 Mining, 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] Sanjeev Dhawan, Kulvinder Singh and Monika, "Critical Investigation on Collaborative Recommendation System Based on User's Credibility", Vivechana: National Conference on Advances in Computer Science and Engineering (ACSE-2016)(Accepted and to be Published in Proceedings) .
- [11] Monika, Sanjeev Dhawan and Kulvinder Singh, "Collaborative Filtering based upon User's Reviews", Advances in Computer Science and Information Technology(ACSIT),ACSIT 2016,p.p. 310-313.