International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 5
May 2018

Movie Recommender System using Item based Collaborative Filtering Technique

[1].Ms.Nethra Sri.,
Mtech.,Asst Prof
Department of Computer Science and Engineering
St.Peter's Engineering College
Hyderabad, India

[2].P.Abhilash [3].K.Avinash , [4].S.Rakesh [5].Ch.Surya Prakash
Department of Computer Science and Engineering
St.Peter's Engineering College Hyderabad, India

Abstract— Recommender systems being a part of information filtering system are used to forecast the bias or ratings the user tend to give for an item. Among different kinds of recommendation approaches, collaborative filtering technique has a very high popularity because of their effectiveness. These traditional collaborative filtering systems can even work very effectively and can produce standard recommendations, even for wide ranging problems. For item based on their neighbor's preferences Collaborative filtering techniques creates better suggestions than others. Whereas other techniques like content based suffers from poor accuracy, scalability, data sparsity and big-error prediction. To find these possibilities we have used item-based collaborative filtering approach. In this Item based collaborative filtering technique we first examine the User item rating matrix and we identify the relationships among various items, and then we use these relationships in order to compute the recommendations for the user.

Keywords— Collaborative filtering Technique; Item based collaborative filtering technique; Recommender Systems; User item rating matrix;

I. INTRODUCTION

The quantity of facts provided or learned about something is increasing very quickly surpassing our ability to process all the information made available. Many of us may be got confused or deluged by the number of movies that are releasing each year or the books that are being released or by the journals that are available for us . It is the right time for us to have some automation in skimming the information that is available for us. Now everything has gone online and because of this connectivity we are made available with loads of information. Because of this loads of information made available for us we may face the situation of selection. We may choose from thousands of movies or lakhs of books and millions of document that are available from various resources. From this numerous choices it may be difficult for us to choose from this numerous choices. One of the solution is to decide ourselves based on past history and our tastes and the next is to seek recommendations from our neighbor's. One of the most promising such technologies is Recommender Systems. These Recommender systems have been a very huge success in giving personalized recommendations to the user. Exploration done in the field of recommender system [1] is going on for several years as of now but the attentiveness in this diversified still Remains high because of the copious amount of the applications as well as the problem space. Huge number of such online applications is made available for us through the amazon site for books and YouTube for Movies and many more.

These Recommender systems are the different kinds of applications that work on by combining the information that has got from the user item rating matrix domain and strains the knowledge to get the most pertinent information. By using this most pertinent information we can calculate the personalized recommendations for the user. From the past few years numerous different types of recommender systems are being introduced by the

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 5
May 2018

researches for different types of problem space. We can calculate these recommendations using recommender systems by different techniques like collaborative filtering technique, Content based technique and hybrid technique which is a mix of different kinds of recommender systems. This recommender system returns the most popular items or movies that customers bought in addition to the selected item or movie at some point in time. In other words: other books or movies customers also purchased/liked.

The first Content based collaborative filtering approach [2], [3] are constructed upon the fact that "Discover the things in the manner of things liked priory". This Recommender system gains information on the preferences of the user through the user ratings. The assessment may be direct in the form of ratings mentioned as like or Dislike by the user for that particular item. Another way of providing he rating is through indirect method in which the user takes the chance to view the movie up on his interest. Where as in Collaborative filtering approach , we first find out the users who are similar to the current user and then calculate the recommendations to the current user. The Three Columns of this approach are many users has to participate in the system and the way through which the users express their preferences must be an effortless way. These Collaborative filtering techniques are used by Bell core Video Recommender systems [5], Group Lens Movie recommender systems [4], and even the Firefly recommender systems [6].

This Collaborative filtering approach is mainly classified into two types they are Model based approach and Neighborhood based approach. The first one neighborhood based collaborative filtering technique approach [7], [8] we will be using the user item rating matrix in order to calculate the ratings that are not rated by the user based up on similar items or users. Hence this finding up of similar users or items can be done in two methods of them the first is Item based collaborative filtering technique and the next one is User based collaborative filtering technique. The first one Item based collaborative filtering approach technique [8], is used for prediction of the unknown ratings for the user for an item based up on the similar items for the item for which we are predicting. The next User based collaborative filtering approach technique is used to calculate the prediction of the unknown ratings for the user for an item based up on the similar users of the user for which we are predicting The opposite for Neighborhood based is the Model based approach. The main theme of this model based approach is to create a model that uses the ratings in the user item rating matrix directly and then instruct the model using the available information and then used for prediction purpose.

In this we will be using the item based collaborative filtering approach technique. The Data set that we have utilized is the Netflix Data set. This Netflix data set is available in the Group Lens which has collected and made accessible of this user item rating data sets from the Movie Lens web site. For the present system we will be using the Stable benchmark data set which consist of around One million ratings from 6040 users on 4000 movies. For calculating the similarities between the items we will use adjusted cosine similarity. Then we use these similarity weights calculated to calculate the predicted rating of the movies or items that are not rated by the user and then give The top most N number of recommendations to the users as recommendations which will be the output.

II. EXISTING WORK

The efforts that are laying down on this Movie recommender systems has been increasing day by day to a greater extent. Not only in movies this recommender system has been used in diversified fields like Books, Documents, publications and many more [1]. The main reason behind this increase in popularity of recommender systems is the competition that has been started by the Netflix organization [5], whose primary motto is to increase the accuracy of the recommendations provided to the user by tem percent.

These Recommendation systems are generally categorized in to two types they are Collaborative filtering approach and Content based approach Techniques. The first one Collaborative filtering approach uses the similarities between users or items that are computed using the user item rating matrix for prediction purpose. Whereas the Content based approach gives recommendations to the user based up on the past history of the user rather than the similar items or users. In model Based collaborative filtering approach we will be using a model which is first trained by the available data and then used for prediction purpose [11]. Off all

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 5
May 2018

these methods Collaborative based item or user filtering approaches are the most popular one because of their efficiency [9].

Tapestry [9] was the first one to use this collaborative filtering techniques to implement recommender systems. In that system the preference of the users are first extracted from the ratings that are given by the user explicitly or implicitly. After this a large number of methodologies has been introduced in order to provide personalized recommendations to the user. Ringo video Recommender system [11] is a web based application that generates recommendations to the user on movies, Videos and music and many more. Group lens [11] also developed a recommender system using item based collaborative filtering approach that provides recommendations for news, Movies etc.

There are many other kind of methodologies that has been introduced in order to implement this Recommendation system which includes diversified fields of Data mining, Clustering, Horting and Bayesian Network Methodology. Off these Bayesian networks works effectively which involves construction of a model and then training the model using the available data and then later used for prediction purpose. Model that is constructed using these Bayesian network methodology works fine, smaller in size and effectively. The main disadvantage of this Bayesian networks is that they cannot be applied for the systems in which the information from which we extract the preferences is frequently changing. In addition to these we have another class of recommender systems in which we will be combining two or more types of recommender systems which are termed as Hybrid Recommender systems which combines the good attributes of different recommender systems there by reducing the anomalies.

Even though these collaborative filtering techniques having huge popularity, Efficiency and wide range of applicability still they face many problems including Cold start problems, Data sparsity and shriller attacks etc., Due to these problems and to improve the performance of CF technique various new approaches have been developed over the years. To solve the problem of sparse user-item matrix various techniques like Singular Value Decomposition [11] and models like Bayesian classifiers, matrix factorization and genetic algorithms are used [12]. But these methods are expensive methods in terms of computations. Various clustering techniques like Particle Swarm Optimization [15], Ant Colony Optimization [15] and k-means [13] have been used to improve the quality of predictions thus provide solutions to remove the cold start problem.

Various trust based recommender systems have been evolved to eradicate the shilling attack and improve the recommendations by incorporating the trust value in the user graph [14]. The recommendations are generated only through trustworthy users. Various trust calculation techniques have been discussed in [15] and it has been clearly indicated that the recommendations always come from the trusted users thereby removing the problem of shilling attack.

These recommender systems can be applied for a wide range of problem domains including books, Electronic media and Entertainment. One who wants to implement this recommender system first they have to understand the end user taste and the preferences of him. The recommender system that we are going to implement should resemble the taste of the user and his requirements and must be suitable for the problem domain. At last the Recommender system that we would like to design should be able to integrate the preferences or the feedback that are provided by the user into a single unit data source so that the recommendations that are calculated are going to represent the entire range of information provided by the user.

In the proposed work, a methodology has been implemented to overcome a few of the above mentioned problems by using adjacent cosine based similarity for computation of similarities and selection of neighbor's and then use them for the prediction ratings. Later recommend items for the user. Advantages of the present system are as follows

1. Improved prediction accuracy when compared to other techniques like content based.
2. It can even work well when we have sparse training data set too.
3. It reduces the number of big error predictions.

III. DATA SET

The Data set that we are using for the present system is the Netflix user item ratings Data set. It is

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 5
May 2018

collected and maintained by Group Lens Research Organization and has collected and made available this user item ratings data set from the Movie Lens web site. The data sets was gathered over various intervals of time. And for our present system we will be using a dataset that consist of one million ratings as preferences that are given by 6040 users over for 3952 Movies.

The ratings that are given by the users as preferences are taken as a single file as ratings.dat file which is available in the Group lens site in the following format as User ID: Movie ID: Rating: timestamp in which the User id will be ranging in between 1 and 6040 Movie IDs range between 1 and 3952 Ratings are made on a 0 to 5 star scale and the Timestamp is used to represent the seconds as the epoch is returned by the time and the user that are represented in the system are going to have minimum of 20 ratings and a maximum of 200 ratings and an average of 40 ratings by user.

These files contain one million ratings that are given by user as preferences for almost 4000 movies made by around 6040 users.

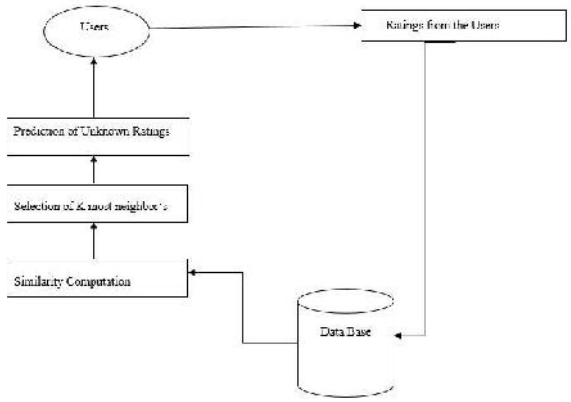

Fig. 1 Architecture of Movie recommender System using Item Based Collaborative Filtering technique.

IV. METHODOLOGY

These Recommendation systems are applicable to a rich numerous number of problem spaces and wide range of applications including books, Movies, Documents and Articles. By using these Recommender systems we can generate personalized recommendations to the user based on his preferences. These Personalized recommendations provide us a great of providing justification for the recommendations that have got generated. Hence in order to satisfy the users that recommendations that are generating should satisfy the users as well as they should be reliable.

In this present system we have detailed the theoretical analysis of the methods that we have utilized for the Implementation of collaborative item based method. In this item based Recommendation process, we generally look at ratings given to similar items. In contrast with the User based Collaborative filtering approach in which we will be looking for the most similar users for the current user in Item based collaborative filtering approach we will be using the items that are most similar to the current item for which we are going to predict the rating by using the item similarity weights and using the K most similar items and predicting the unknown rating. Then we will recommend the top N items having highest predicted rating as recommendations to the user.

A. Computation of Similarity Weight

This similarity weight is going to play an important role in the collaborative item based filtering approach and hence in order to maintain or select the trustable users from the given set of user. Hence they give us a method to increase or decrease the significance of a particular user or item. In the present methodology we are using adjusted cosine similarity for computation of similar weights of items.

$$AC(i,j) = \frac{\sum_{U \in u_{ij}} (r_{ui} - \bar{r}_u)(r_{uj} - \bar{r}_u)}{\sqrt{\sum_{U \in u_{ij}} (r_{ui} - \bar{r}_u)^2 \sum_{U \in u_{ij}} (r_{uj} - \bar{r}_u)^2}}$$

Where in this r_{ui} representing the rating that is given by the user u to the item i and r_{uj} represents the rating that is given by user u to item j, $\bar{r}_u$ indicates the average ratings that are given by user u as a whole in the rating matrix.

B. Selection of Neighborhood

In this Collaborative filtering approach the number of neighbors that we are going to use as a part of prediction also creates a significant impact on the quality of recommendations that are going to be generated. Hence these selection of Neighbors has to be done more carefully so as to not affect the

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 5
May 2018

quality of recommendations generated. Hence we will be choosing the K most similar neighbors which are having the highest similarity compared to others. So this value of K must be chosen more carefully.

C. Prediction of Unknown Ratings

In this for the current user for whom we are intended to give predictions those items for which the user hasn't rated should be predicted using the similar weights and selecting the K most similar weights that is the K^{th} most similar items that we have computed in the previous step are used for the predictions of unknown rating and it is calculated using the following formulae

$$r_{ui} = \frac{\sum_{j \in N_u(i)} W_{ij} \, r_{uj}}{\sum_{j \in N_u(i)} | W_{ij} |}$$

Where W_{ij} represents similarity weight between items i and j. r_{uj} Will representing the rating that is given by the user u to the item j. N_u (i) represents users that have rated item i.

D. Recommending Top N Items

In this process out of the predicted values that are not rated by the current user the top N items which are having highest predicted value are given as recommendations to the user. Value of N should be selected carefully so as to give proper recommendations for the user.

V. EVALUATION METRICS

Precision is one of the important measure that is used in order to evaluate the accuracy of recommendations that are generated. Generally the User rating data set that is available in the Group lens that we are utilizing is taken and it is divided into two sets and one of the set is termed as R_{train} which is used to train the Algorithm and used to learn and the next set is termed as the test set R_{test} which is used to evaluate the accuracy of predictions generated. One of the important technique that is used to analyse and measure the accuracy and precision of the Recommendations generated is the Mean absolute Method termed as MAE in Acronym. Mena absolute error which is termed as MAE is defined as the measure of deviation or divergence of the predicted ratings through content based collaborative filtering technique from the original ratings. It is calculated as the mean or average of the absolute errors that are calculated and it can be defined as in the following manner:

$$MAE(f) = \frac{1}{|R_{test}|} \sum_{r_{ui} \in R_{test}} (f(u,i) - r_{ui})$$

Where R_{test} represents the training set and r_{ui} represents the ratings that are given by the user u to the item i, and f (u, i) represents the actual rating that are given by the user u to the item i in the test set that we have taken. A lower Mean Absolute Error value indicates that the recommendations that are generated by the present system are accurate. So generally smaller mean absolute errors are generally recommended.

A. Impact of number of neighbors used

In order to test the influence of the number of neighbors used in order to calculate the recommendations generally we vary the number of neighbors that is the value of K used from ranging from 10 to 10 and the mean absolute error is calculated. Although the Mean Absolute Error values for some values of K e.g., K = 30 are a little bit higher than those for other values of K e.g., K = 20. Thus, we maintain the quality of recommendations by selecting a suitable threshold value of K.

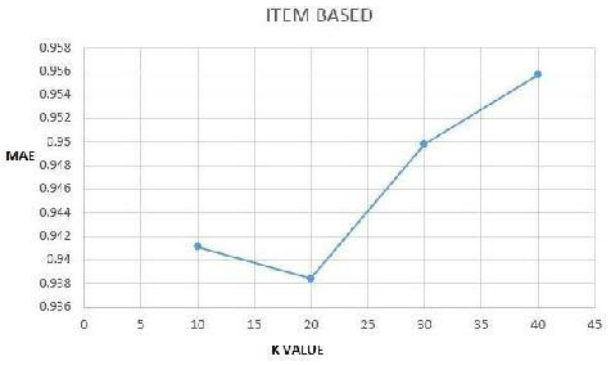

Fig. 2 Plotting of MAE for Different Values of K using item based collaborative filtering technique.

Based on various MAE Values for various values of K we find that the MAE is low at when K=20 and it is getting increased consistently after K=20 and increasing gradually till 40 which is our range of the K.

VI. CONCLUSION

Recommendations that are generated using Item based collaborative filtering technique are easy to implement, Reliable and Justifiable. In the system it is better to use Item based approach if Users are far greater than the number of items. The performance of this collaborative filtering approach is effected by data sparsity, cold start problem and shriller attacks for new users and hence there is a great chance of conducting in this area. As the need for this Recommender systems is increasing drastically new technologies are needed to increase its performance.

In the present paper we have evaluated Collaborative item based approach and evaluated the recommendations for the current user. Our results hold the promise of using Collaborative filtering approach even for large scale data.

References

[1] P. Resnick and H. R. Varian: "*Recommender Systems*", Communications of the ACM, vo1.40, pp.56-58, 1997

[2] Lieberman, H.: "*Autonomous Interface Agents*", in Proceedings of CHI'97 (Atlanta GA, March 1997), ACM Press, 67-74.

[3] Maes, P.: "*Agents That Reduce Work and Information Overload*", Communications of the ACM 37, 7, 31-40, July 1994.

[4] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, and J. GroupLens: "*An Open Architecture for Collaborative Filtering of Netnews*". In Proceedings of CSCW'94 (Chapel Hill NC, October 1994), ACM Press, 175-186

[5] Hill, W.C., Stead, L., Rosenstein, M. and Furnas, G. "*Recommending and Evaluating Choices in a Virtual Community of Use*", in Proceedings of CHI'95 (Denver CO, May 1995), ACM Press, 194-201.

[6] Shardanand, U., and Maes, P. "*Social Information Filtering*": *Algorithms for Automating* "Word of Mouth". In Proceedings of CHI'95 (Denver CO, May 1995), ACM Press, 210-217.

[7] Delgado, J., Ishii, and N.: "*Memory-based weighted majority prediction for recommender systems*". In: Proc. of the ACM SIGIR'99 Workshop on Recommender Systems (1999)

[8] Deshpande, M., Karypis, G.: "*Item-based top-N recommendation algorithms*". ACM Transaction on Information Systems 22(1), 143– 177 (2004)

[9] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). "*Using Collaborative Filtering to Weave an Information Tapestry*". Communications of the ACM. December.

[10] Konstan, J., Miller, B., Maltz, D., Herlocker, J.Gordon, L., and Riedl, J. (1997). Group Lens: "*Applying Collaborative Filtering to Usenet News*". Communications of the ACM, 40(3), pp. 77-87.

[11] A. B. Barragáns-Martínez, E. Costa-Montenegro, J. C. Burguillo, M. Rey-López, F. a. Mikic-Fonte, and A. Peleteiro, "*A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition*," Inf. Sci. (Ny)., vol. 180, no. 22, pp. 4290–4311, Nov. 2010.

[12] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "*Knowledge Based Systems Recommender systems survey*," Knowledge-Based Syst., vol. 46, pp. 109–132, 2013.

[13] G. M. Dakhel, "*A New Collaborative Filtering Algorithm Using K-means Clustering and Neighbors Voting*," pp. 179–184, 2011.

[14] J. Sobecki, "*Ant Colony Metaphor Applied in User Interface Recommendation*," vol. 26, pp. 277–293, 2008

[15] S. Alam, G. Dobbie, P. Riddle, and Y. S. Koh, "*Hierarchical PSO clustering based recommender system*," 2012 IEEE Congr. Evol. Comput pp. 1–8, Jun. 2012

Author Bibliography

P.NethraSri, Asst.Professor, Dept.of CSE, St.Peters Engineering college, Hyderabad, TS, INDIA (nethra.sri08@gmail.com)

Pothineni Nethrasri received M.Tech in computer science and Engineering from Jawaharlal Nehru technological university, Hyderabad. She is currently working As an Assistant Professor, Department of Computer Science and Engineering at St. Peters Engineering College, Hyderabad, TS, INDIA. She has published papers at International Journal of Innovative Technology and Research (IJITR) and in the National conference on recent trends in Big Data Analytics and cloud computing.(NCRTBDACC).