International Journal of Innovative Research in Engineering
Volume 3, Issue 3 (May-June 2022), PP: 628-631.
www.theijire.com
ISSN No: 2582-8746

Cine hunt: Movie Recommendation Engine

Aneez Rahman[1], Edwin Saju[2], E S Amal Akhtar[3], Neetha K Nataraj[4]
[1,2,3,4] *Dept. of CSE, Adi Shankara Institute of Engineering and Technology, Kalady, Kerala, India.*

How to cite this paper:
Aneez Rahman[1], Edwin Saju[2], E S Amal Akhtar[3], Neetha K Nataraj[4], "Cine hunt : Movie Recommendation Engine", IJIRE-V3I03-628-631.

Copyright © 2022 by author(s) and5th Dimension Research Publication.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Abstract— *The main objective of the project is develop a web application which recommends movies using user based collaborative filtering. Recommendation system produce a ranked list of items on which the user might be interested .Helps in deciding which movie to see the next Depends on users rating and likes and dislikes Can be contructed in collaborative filtering , content based approach and hybrid models. Content based – It recommends other movies which are similar to selected user .Collaborative filtering – It recommends movies which are rated highly by the similar users.*

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

I. INTRODUCTION

A movie recommendation system , is a recommender system which is used to suggest movies to the user . It produces a ranked list of items on which the user might be interested. The information can be collected in two ways, implicit and explicit. An implicit acquisition involves observing the user's behavior such as watched movies on the other hand explicit acquisition involves collecting the users previous rating or history.

Collaborative Filtering is a way of filtering or calculating items through the sentiments of other people . It first gathers the user ratings given by individuals and then recommends movies to the target users based on likeminded people with similar tastes and interest in the past.

Movies are being loved mostly by everyone irrespective of their age ,colour ,location or race rather they have their own choice of genres like thriller ,action ,drama etc and some like movies because of some actor or actress . For lessen the work of users i.e. for searching movies of specific genres or actor / actress ,a recommender system is prescribed to work in background .A movie recommender system is such which suggest movies according to the users like and ratings . For example if user 1 has watched three movies where as user 2 has watched only first and third movie and user 3 has watched first and second movie ,so movie 2 will be recommended to the user 2 and movie 3 to the user 3 as they have watched the other two.

II. EXISTING WORK

The significance of recommender systems has been increasing day by day. Not only in movies this recommender system has been utilized in diversified fields like Books, Documents, publications and lots of more [19]. The most reason behind this increase in popularity of recommender systems is that the competition that has been started by the Netflix organization [20] [1], whose primary motto is to extend the accuracy of the recommendations provided to the user by one-tenth.

These Recommendation systems are often classified into two types. They're Collaborative filtering approach and Content based approach Techniques. The primary 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 au courant the past history of the user instead of the similar items or users. In model Based collaborative filtering approach we first train the model by the available data and may be used for prediction purpose. Off of these methods Collaborative based item or user filtering approaches are the foremost popular one due to their efficiency.

Tapestry was the primary one to use these collaborative filtering techniques to implement recommender systems. In this system the preference of the users are first extracted from the ratings that are given by the user explicitly or implicitly. After this an outsized number of methodologies has been introduced so as to supply personalized recommendations to the user. Ringo video Recommender system may be a web based application that generates recommendations to the user on movies, Videos and music and plenty of more. Group lens also developed a recommender system using item based collaborative filtering approach that has recommendations for news, items, Movies etc.

There are such a large amount of other different techniques that are developed to implement a recommendation system. These techniques include various diverse fields of clustering [2][1] , data processing and Bayesian Network Methodology. These methodologies work effectively that construct a model using great amount of knowledge that's training data and so use this model for the prediction of the output for brand spanking new input. Model, constructed using these methodologies, works fine and effectively.

III. RELATED WORK

A. Memory-based Algorithms

Memory-based algorithms, also referred to as neighbour based algorithms, operate entire database of ratings collected by the seller or service supplier. The Memory-based algorithms are widely employed in many large commercial sites, like Amazon etc. At present, there are many ongoing researches focused on developing highly reliable Memory-based algorithms. Most of the researches improve the accuracy of Memory-based algorithms only by improving the similarity measures. Usually there are two models to live the similarity of users. They're Pearson coefficient of correlation (PCC) [4] and Vector similarity (VS) [5]. PCC and VS are very simple, but they both have a shortcoming which only consider the co-rated items. It may lead to a controversy that two users may have a high similarity only because they need few co-rated items and coincidently ranked this stuff similarity. Therefore, Hao Ma et al, proposed to feature a correlation significance weighting factor that may evalue similarity weights that were supported atiny low number of corated items [6]. Additionally to the above methods, reference [7] [8] also proposed similarity measures by using the graph theory. Even more, Heng Luo et al, proposed a collaborative filtering framework supported both of local user similarity and global user similarity [9]. The above research about the similarity measure does improve the accuracy of the Memory-based algorithms. But few researches focused on the prediction score models which we believe are more important than the similarity measure. During this paper, our study will fill the gap.

B. Model-based Algorithms

Model-based algorithms are different with Memory-based algorithms. It first uses the database to estimate or learn a model then apply this model for prediction. Generally speaking, the Model-based algorithms usually have higher accuracy than the Memory-based algorithms. Among the model-based algorithms, the foremost representative is that the matrix factorization. Over the hobbies, plenty of matrix factorization techniques are proposed, including singular value ecomposition [10], probabilistic latent semantic analysis [11], probabilistic matrix factorization [12] and etc. Taking into consideration the non-public difference, reference [13] [14] proposed a bias feature idea. However, the algorithms mentioned above only consider the educational process, and ignoring the adjustment process after training. during this paper, we proposed a differential model which might be applied to any matrix factorization.

C. Hybrid Algorithms

Hybrid recommender systems by combining each strategy together can provide better performance instead of either strategy alone [15] [16]. The foremost famous is that the BellKor's solution winning the Netflix prize [17], which mixes predictions from 107 different baseline recommender systems. By Burke's survey [18], the hybrid recommender systems are mainly divided into the subsequent classes. They're mixed, switched, weighted, feature argumentation and meta-level hybrids. During this paper, we proposed a weighted hybrid methods which might avoid or catch up on the shortcomings of matrix factorization and neighbor-based methods. The remainder of the paper is organized as follows. The subsequent section provides an outline of the differential model proposed by us. Section 4 provides an outline of the improved Neighbor-based methods proposed by us. In section 5, we describe the hybrid method thoroughly. Section 6 provides the experimental results. Finally, conclusions and future works are provided in section 7.

IV. DATASET

The Data set that we are using for the current system is that the Netflix user item ratings Data set. it's collected and maintained by Group Lens Research Organization and has collected and made available this user item ratings data set from the Movie Lens site. The info sets was gathered over various intervals of your time. And for our present system we are going to be employing a dataset that carries with it 1,000,000 ratings as preferences that are given by 6040 users over for 3952Movies.

The ratings that are given by the users as preferences are taken as one file as ratings.dat file which is on the market within the Group lens site within the following format as User ID: Movie ID: Rating: timestamp during which the User id are {going to be|are} ranging in between 1 and 6040 Movie IDs [1] range between 1 and 3952 Ratings are made on a 0 to five star scale and also the Timestamp is employed to represent the seconds because the epoch is returned by the time and therefore the user that are represented within the system are going to have minimum of 20 ratings and a maximum of 200 ratings and a mean of 40 ratings by user. These files contain 1,000,000 ratings that are given by user as preferences for pretty much 4000 movies made by around 6040 users.

V. METHODOLOGY

These Recommendation systems are applicable to a chic numerous number of problem spaces and wide selection of applications including books, Movies, Documents and Articles. By using these Recommender systems we will generate personalized recommendations to the user supported his preferences [1]. These Personalized recommendations provide us a good of providing justification for the recommendations that have gotten generated. Hence so as to satisfy the users that recommendations that are generating should satisfy the users moreover as they ought to be reliable. In this present system we've got detailed the theoretical analysis of the methods that we've got utilized for the Implementation of collaborative item based method. During this item based Recommendation process, we generally have a look at ratings given to similar items. In contrast with the User based Collaborative filtering approach within which we are {going to} be searching for the foremost similar users for this user in Item based collaborative filtering approach we'll be using the things that are most kind of like the present item that 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'll recommend the highest N items having highest predicted rating as recommendations to the user.

a. **Computation of Similarity Weight**

This similarity weight goes to play a vital role within the collaborative item based filtering approach and hence so as to take care of or select the trustable users from the given set of user. Hence they furnish us a way to extend or decrease the importance of a selected user or item. Within the present methodology we are using adjusted cosine similarity for computation of comparable weights of things.

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

b. **Selection of Neighborhood**

In this Collaborative filtering approach the quantity of neighbors that we are visiting use as part of prediction also creates a major impact on the standard of recommendations that are visiting be generated. Hence these selection of Neighbors needs to be done more carefully so on not affect the standard of recommendations generated. Hence we are going to be choosing the K most similar neighbors which are having the best similarity compared to others. So this value of K must be chosen more carefully

c. **Prediction of Unknown Ratings**

In this for this user for whom we are intended to administer predictions those items that the user hasn't rated should be predicted using the similar weights and selecting the K most similar weights that's the Kth most similar items that we've computed within the previous step are used for the predictions of unknown rating and it's calculated using the subsequent formulae

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

d. **Recommending Top N Items**

In this process out of the expected values that aren't rated by this user the highest N items which are having highest predicted value are given as recommendations to the user [1]. Value of N should be selected carefully soon give proper recommendations for the user.

VI. 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 Rtrain which is used to train the Algorithm and used to learn and the next set is termed as the test set Rtest 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. Mean 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 Rtest represents the training set and Rw 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 [1]. 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.

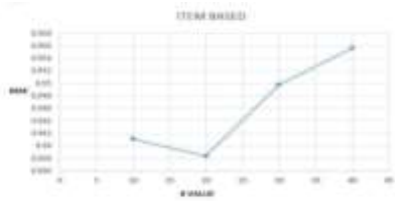

Fig. 1 Plotting of MAE for Different Values of K using item based collaborative filtering technique.

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.

VII. CONCLUSION

Recommendations that are generated using Item based collaborative filtering technique are easy to implement, Reliable and Justifiable. within the system it's better to use Item based approach if Users are far greater than the quantity of things. The performance of this collaborative filtering approach is effected by data sparsity [1], cold start problem and shriller attacks for brand spanking new users and hence there's a good chance of conducting during this area. because the need for this Recommender systems is increasing drastically new technologies are needed to extend its performance. within the present paper we've got evaluated Collaborative item based approach and evaluated the recommendations for the present user. Our results hold the promise of using Collaborative filtering approach even for big scale data.

The work, item-based movie recommendation system has been described which uses the rating of flicks as its feature. It uses the modified cosine similarity matrix to seek out movies needed to recommend. [2]For the give dataset, the User-based collaborative filtering outperformed Item-based collaborative filtering.

References

[1] Lakshmi Tharun Ponnam, Sreenivasa Deepak, Siva Nagaraju and Srikanth Yellamati "Movie Recommender System Using Item Based Collaborative Filtering Technique" IEEE 2016 International conference on Emerging trends in Engineering , Technology and Science.
[2] Mukesh Kumar Karita, Atul Kumar and Pardeep Singh "Item-Based Collaborative Filtering in Movie Recommendation in Real time" IEEE 2018 First International Conference on Secure Cyber computing and Communication, 2018.
[3] Rui-sheng Zhang, Qi-dong Liu, Chun-Gui, Jia-Xuan Wei, Huiyi-Ma "Collaborative Filtering for Recommender Systems" IEEE 2014 Second International Conference on Advanced Cloud And Big Data.
[4] Resnick P, Iacovou N, Suchak M, et al. GroupLens: an open architecture for collaborative filtering of netnews[C]//Proceedings of the 1994 ACM conference on Computer supported cooperative work. ACM, 1994: 175-186
[5] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In Proc. of UAI-98, 14th Conf. on Uncertainty in Artificial Intelligence, pages 43–52, Madison, Wisconsin, USA, 1998.
[6] Ma H, King I, Lyu M R. Effective missing data prediction for collaborative filtering[C]//Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2007: 39-46.
[7] Gori M, Pucci A. ItemRank: A Random-Walk Based Scoring Algorithm for Recommender Engines[C]//IJCAI. 2007, 7: 2766-2771.
[8] Fouss F, Pirotte A, Renders J M, et al. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation[J]. Knowledge and Data Engineering, IEEE Transactions on, 2007, 19(3): 355-369.
[9] Luo H, Niu C, Shen R, et al. A collaborative filtering framework based on both local user similarity and global user similarity [J]. Machine Learning, 2008, 72(3): 231-245.
[10] M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. Application of dimensionality reduction in recommender system—a case study. In Proc. of WebKDD-00, Web Mining for E-Commerce Workshop, at 6th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Boston, Massachusetts, USA, 2000.
[11] Hofmann. Latent semantic models for collaborative filtering. ACM Trans. Inf. Syst., 22(1):89–115, 2004.
[12] Salakhutdinov and A.Mnih. Probabilistic matrix factorization. In J. C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, Advances in Neural Information Processing Systems 20.MIT Press, Cambridge, Massachusetts, USA, 2008.
[13] Paterek, Arkadiusz. "Improving regularized singular value decomposition for collaborative filtering." Proceedings of KDD cup and workshop. Vol. 2007. 2007.
[14] Takacs, Gabor, et al. "On the gravity recommendation system." Proceedings of KDD cup and workshop. Vol. 2007. 2007.
[15] Yuan-hong, Wu, and Tan Xiao-qiu. "A real-time recommender system based on hybrid collaborative filtering." Computer Science and Education (ICCSE), 2010 5th International Conference on. IEEE, 2010.
[16] Xu, Hai-Ling, et al. "Comparison study of Internet recommendation system." Journal of software 20.2 (2009): 350-362.
[17] Bell, Robert M., Yehuda Koren, and Chris Volinsky. "The BellKor solution to the Netflix prize." KorBell Team's Report to Netflix (2007).
[18] Burke, Robin. "Hybrid recommender systems: Survey and experiments." User modeling and user-adapted interaction 12.4 (2002): 331-370
[19] P. Resnick and H. R. Varian: "Recommender Systems", Communications of the ACM, vo1.40, pp.56-58, 1997
[20] 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