A Hybrid Neighborhood based Book Recommendations System

Pooja B. Jire
M. E. Student, Computer Department,
R. C. Patel Institute of Technology,
Shirpur, Dist.-Dhule, (M. S.)

Prof. R. B. Wagh
Assistant Prof., Computer Department,
R. C. Patel Institute of Technology,
Shirpur, Dist.-Dhule, (M. S.)

Abstract: **The overload of data and excessive available information on web triggered the new research area of recommender systems. The recommender system (RS) is a way to provide the information which user may be interested in while surfing the web. RS can personalize and customize the Web environment in real-time. There are many real time examples such as flipkart.com, amazon.com and many e-commerce websites. There are multiple surveys and research happened in past in this area and various techniques were presented to provide better recommendation to web users. In this paper, we have analyzed and implemented the recommender system based on hybrid algorithm which uses the user to user, item to item and slope one algorithm. Recommendations are provided to users by calculating the heat values between items and users. We perform experimental comparison of the hybrid algorithm based method against other individual recommendation algorithms with Book Crossing real data set. Results shows significant similarities and slight improvement in terms of effectiveness and accuracy measured through RMSE method.**

Keywords—Recommender System, Slope One, Hybrid recommendation, RMSE.

I. INTRODUCTION

A few years ago, there was an Age of T.V. and everybody was crazy about it. In the same way current age we can call an age of Internet. As a minimum one person in a home is using it. Because of which it also creates the new challenges in the world of web browsing and that's where the Web Recommendation plays the vital role in the internet world. Examples of such recommender systems are recommendations for books, music, news, jokes and web pages.

The web is growing explosively and may study shows that the web pages are increased exponentially every year. Finding desired information is more difficult for users when surfing on an internet. Users often wander aimless on the Web without visiting pages of his/her interests and spend a long time on finding the expected information.

The main objective of web page recommendation is to understand users' navigation behavior and to show some pages of users' interests at a specific time. There exists lot of existing models on this topic such as Markov model and its variants and temporal relation is important.

New methods of development for online information services are required now because of growth in use and size of World Wide Web. Most of the web structures are complicated and not user friendly plus larger in size. Hence users often fail to achieve the goal of their required search and receive irrelevant results which further change the direction of search. Therefore it is required to predict the user needs to improve browsing experience and providing them with what they want.

Web recommendation systems have become popular research and development topic as its help web users getting information they are interested in. Generally web recommendation system expects user's inputs explicitly.

Web-page recommendation is to predict the next request of pages that users are possibly interested in when surfing the internet. This technique can help Web users to locate more useful pages without asking for them clearly and has attracted much attention in the community of Web mining [4].

A. Collaborative Filtering

Collaborative Filtering (CF) is perhaps the most studied and also the most widely-used recommendation approach in practice. The key characteristics of CF are it predicts the utility of items for a user based on the items previously rated by other like-minded users. The main idea of CF is People who liked Book1 may also like Book2 if user rated Book1 and Book2 in the similar fashion.

Collaborative methods overcome some of the drawbacks of content-based ones. For instance, items for which the content is not available or difficult to obtain can still be recommended to users through the feedback of other users. Furthermore, collaborative recommendations are based on the quality of items as evaluated by peers, instead of depending on content as that can decrease the quality of results.

The main reason behind using collaborative is because some people have gone to the effort of viewing/filtering things, and chosen the best from the huge information. Other users get a recommendation of the best few, without having to spend the same effort again. Some very good examples of CF: twitter, PageRank, Facebook (Likes), etc.Many systems ask users to rate items – e.g. on a scale of1 to 10. These ratings then enable the system to give moreprecise/accurate recommendations, and use a variety of sophisticated learning/prediction algorithms.

Hao Ma et al. [1] proposed a Graph based collaborative filtering techniques for recommendation. The simplest and original implementation of this approach recommends to the active user the items that other users with similar tastes liked in the past. The similarity in likelihood of two users is calculated based on the similarity in the rating history of the users. This is the reason why refers to collaborative filtering as "people-to-people correlation".

Figure 1 depicts the different techniques of collaborative filtering. In this paper, we have focused more on memory based CF which is also called neighborhood based collaborative filtering.

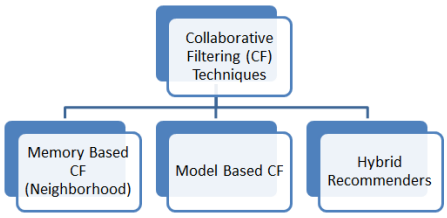

Fig 1.Different techniques of Collaborative Filtering

In today's world e-commerce websites use a recommendation algorithm which provides an effective form of targeted marketing by creating a personalized shopping experience for each customer.

B. Neighborhood Recommendation System

The neighborhood-based CF algorithm, a established memory-based CF algorithm, calculate the similarity between users and items and generates prediction for the current user by taking the average of all the ratings of the user or item on a certain item or user. In this types of recommendation happens based on similar users or similar items.

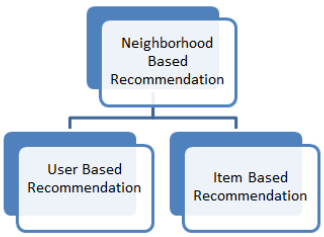

Fig. 2.Different methods of Neighborhood approach

C. Pearson Algorithm

Computing the hyper geometric function pFq is an important problem due to its wide variety of applications in problems in mathematical and theoretical physics, networks and many other areas. Pearson's correlation coefficient between two variables is defined as the covariance of the two variables divided by the product of their standard deviations. It used an item based collaborative filtering technique and determines the similarity between the two items rated together by users. Pearson correlation suffers from computing high similarity between users with few ratings in common. This can be alleviated by setting a threshold on the number of co-rated items necessary for full agreement (correlation of 1) and scaling the similarity when the number of co-rated items falls below this threshold [7].

D. Slope One Algorithm

Many technologies have been developed for recommendation systems. Collaborative filtering (CF) approaches are the most popular and efficient. CF approaches can be grouped into two categories: memory-based and model-based. Empirically, model-based methods may achieve higher prediction accuracy than memory-based methods. While memory-based methods provide a concise and intuitive justification for the computed predictions. Due to the no-paraphrase of model-based method, we do not directly exploit the model-based approaches [9].

Slope one is the simplest form of non-trivial item-based collaborative filtering based on ratings. They can be implemented easily due to their simplicity while their accuracy is often on par with more complicated and computationally expensive algorithms.

II. RELATED WORK

Recommender systems are generally used in many application settings to suggest products, services, and information items to likely users. Recommendation algorithm takes user/item attributes and user-item interactions (ratings, article browsing activities, book borrowing activities, etc.) as input to predict the unseen level of match of a specific user-item pair. Collaborative filtering has been one of the most successful and well-studied recommendation algorithm, which depend on the user-item interaction data to make recommendations. A user-based collaborative filtering algorithm first develops user neighborhoods by finding similar users based on their over lapping interactions or similar ratings of common items. Then it makes recommendations based on a user's neighbors' experiences [6].

Adomavicius G. and Tuzhilin A. have proposed User-based Collaborative filtering (CF) is the most successful system for building recommender systems to date, and is widely used in many e-commerce recommender systems. These schemes depend on the fact that each person belongs to a larger group of similarly-behaving individuals. Accordingly, items frequently purchased by the various members of the group can be used to form the basis of the recommended items [8].

III. PROBLEM DEFINITION

For getting efficient recommendations to the customers from a huge data set is a rising problem to e-shopping. Data sorting from a huge data set is a challenge with proper reliable computation time and efficient separation. In order to derive a formula and representation of recommendation, let us first define some naming conventions. Users accessing the system are a set of users U and items or products can be named as a set of items. Moreover, we represent by R the set of ratings logged and create S as the set of likely values for a rating (e.g., S = [1,2,3,4,5] or S = {Agree, disagree}). Also, we assume that no more than one rating can be made by any user $u \in U$ for a particular item $i \in I$ and write this rating. However, auser needs to provide 3 ratings for items in order to get recommendation from system.

IV. PROPOSED SYSTEM

For getting Statistical recommendations from a data set as per the ratings given by other users to items we proposed a hybrid Algorithm for recommended data. We propagate results in the form of highly rated items from a data set of books. For this Hybrid Algorithm is being used which is the combination of Item to Item for Calculating the similarity, Pearson Algorithm for Person to person calculations based on priority given to the data items and Slope one Algorithm for Rating Predictions. RMSE is a technic which is used for calculating actual errors while getting the recommendations from data set. For this following steps can be followed,

Phase 1:User Authentication and Registration for New User

Phase 2 : Recommendation Choice

a. Through Pearson Method

b. Through Item - to Item Method

c. Through Slope One Method

Phase 3 : Hybrid Method Recommendations

Phase 4 : Comparison to the Best Solution.

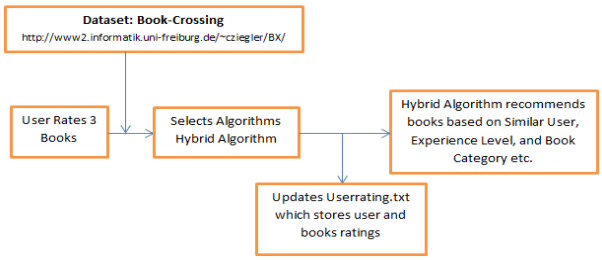

Fig. 3. Data Flow Diagram

A. *RMSE*

For calculating the errors in the recommendations, Root Mean Square Error method is implemented. RMSE is a related measure that has the effect of placing greater emphasis on large errors. By using RMSE, Test Dataset can be simulated and the results of the all recommendations algorithm can be compared. By comparing Hybrid Recommender with Pearson Algorithm and Item to Item Algorithm, conclusion can be drawn as "Our Hybrid Algorithm is least Error Recommender system" [10].

V. IMPLEMENTATION

Web recommendation is to predict the next request of pages that Web users are actually interested in surfing the Web. This system can guide Web users to find more beneficial books (pages) without asking for them explicitly and has attracted much attention in the community of Web mining.

By implementing Pearson algorithm, Item to Item based Algorithm and Neighborhood Recommendation Algorithm we had compare the results of Hybrid Approach to prove that Hybrid Approach can implement to get Least Error recommendation system. Following is the Algorithm for Hybrid Approach.

P={p1,p2,p3…..pn} are set of User

I={i1,i2,i3……in} are set of item

Sp={Sp1,Sp2…..Spn} similarity between current person to other user

I={I12,I13…Inn} Similarity between items

1. Rate any three items
2. Pearson algorithm apply
   a. Similarity between current user and others user are calculated i.eSp
   b. User Rating Selected if(Spn>0.60) then record that rating
3. Item to Item Algorithm Apply
   a. Similarity Between All Items in Dataset are calculated i.e. I
   b. Current user most Popular Item Selected
      i. If(i>3)then record this Item
      ii. Find all most relevant item to record items if(i>0.60) and store in Database
4. Hybrid Recommendation
   a. If rating of Current item is missing then
      i. Difference between current item to others are calculate
      ii. For same user other item rating with difference are added
      iii. Difference is divide by mean value.
      iv. Rating for current user and missing value is predicted.
      v. Pearson and Item to Item ,Most relevant item are recommended.
5. Exit

After the implementation of Pearson and Item to item recommendation algorithm it can get compared using RMSE in the following manner. This propagates the errors of all Algorithms. So, it can be stated as Hybrid Approach can reduce the Errors in Book recommendation system.

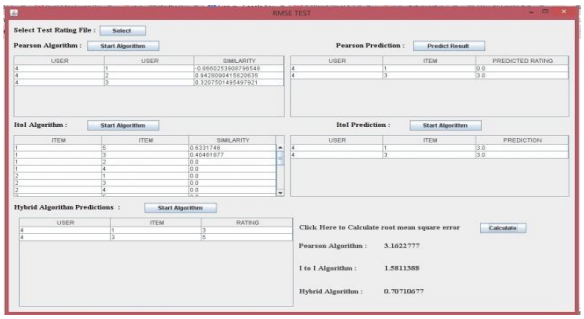

Fig. 4. Implementation of Recommendation Algorithms

The Graphical representation of the comparison between Pearson, Item to Item and Hybrid approach can be drawn below.

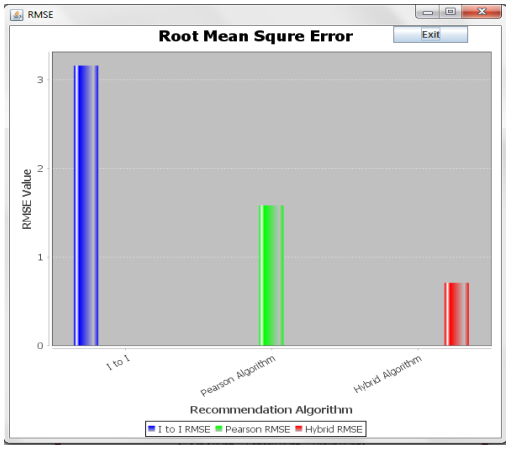

Fig. 5. Comparison using RMSE

From the above figure, Item to Item recommendation algorithm is Most Error algorithm as compare to the Pearson algorithm while Hybrid approach is the Least Error system.

VI. CONCLUSION

Providing meaningful Recommendations to user is very challenging problem due to lack of user's information. Recommendations based on user to user technique becomes challenging when number of users grows and same goes in reverse for item to item based recommendation technique. Hence, we have implemented a recommendation system based on hybrid algorithm which is a combination of user to user , item to item and slope one algorithm. This system takes user ratings as an input and calculates heat values and provides

recommendations considering users experience level, based on hybrid algorithm, category wise. We have used Book Crossing dataset for this system.

We have implemented three ways to get recommendations for books using Pearson algorithm, Item to Item algorithm and Hybrid algorithm.

Item to Item algorithm and Pearson algorithm takes huge time to execute on a large dataset compared to hybrid algorithm which takes output of individual algorithms as an input.

By comparing Hybrid Algorithm with Pearson and Slope One algorithms through RMSE, conclusion can be drawn as Hybrid Algorithm can give Least Error Recommendation.

VII. FUTURE WORK

We can improve the performance of recommendation system by supplying it Recommendation/predicted item result.Find Better hybridization to reduce Root Mean Square Error value. Memory Based Recommendation algorithm is helpful to get better hybrid recommendation algorithms.Memory Based Recommendation algorithm is helpful to get better hybrid recommendation algorithms.

REFERENCES

[1] Hao Ma, Irwin King and Michael Rung-TsongLyu, "Mining Web Graphs for Recommendations", IEEE Transactions on Knowledge and Data Engineering, Vol. 24, No. 6, June 2012.

[2] Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., Riedl, J., "GroupLens: applying collaborative filtering to usenet news", Communications of the ACM 40(3), 77–87 (1997).

[3] Terveen, L., Hill, W., Amento, B., McDonald, D., Creter, J., "PHOAKS: a system for sharing recommendations", Communications of the ACM 40(3), 59–62 (1997).

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

[5] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J., "GroupLens: An open architecture for collaborative filtering of netnews". In: CSCW '94: Proc. of the 1994 ACM Conf. on Computer Supported Cooperative Work, pp. 175–186. ACM, New York, NY, USA (1994).

[6] Huang, Z., Chen, H., Zeng, D., "Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering", ACM Transactions on Information Systems 22(1), 116–142 (2004).

[7] Luo, H., Niu, C., Shen, R., Ullrich, C., "A collaborative filtering framework based on both local user similarity and global user similarity". Machine Learning 72(3), 231–245 (2008).

[8] Adomavicius, G., Tuzhilin, A., "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions", IEEE Transactions on Knowledge and Data Engineering 17(6), 734–749 (2005).

[9] Yongqiang Wang, Liang Yin, Bing Cheng, and Yong Yu, "Learning to Recommend Based on Slope One Strategy", IEEE.

[10] Francesco Ricci, LiorRokach and BrachaShapira,ChristianDesrosiers and George Karypis,Chapter-19, Recommender Systems Handbook