Online Book Recommendation System by using Collaborative filtering and Association Mining

Suhasini Parvatikar
M.E (pursuing)
Computer Engineering Department
Saraswati College of Engineering
Kharghar, Mumbai, India
er.suha@gmail.com

Dr. Bharti Joshi
Ph.D Scholar
Computer Engineering Department
Saraswati College of Engineering
Kharghar, Mumbai, India
bhartijoshi1@rediffmail.com

Abstract – **Recommendation systems is used for the purpose of suggesting items to purchase or to see. They direct users towards those items which can meet their needs through cutting down large database of Information. A various techniques have been introduced for recommending items i.e. content, collaborative and association mining techniques are used. This paper solves the problem of data sparsity problem by combining the collaborative-based filtering and association rule mining to achieve better performance. The results obtained are demonstrated and the proposed recommendation algorithms perform better and solve the challenges such as data sparsity and scalability.**

Keywords-Collaborative filtering, Association rule mining.

I. INTRODUCTION

Recommendation system [11] is the type of information filtering which involves prediction of rating and user preferences, which would help user to buy items according to their needs and interest. The books suggestion at Amazon.com is best example of recommendation system .Recommendation System directs [11] the way to find products, Information according to their interest. Recommendation system uses following technologies to recommend products: Content filtering, Collaborative filtering & Association mining. Content filtering recommends item which is based on Users profile, which he has liked in past. Collaborative based filtering is method to analyze the user's behavior by predicting the users taste to that of similar to other user. Association mining serves as finding relations/correlation among items in a large database. An association rule is a condition of the form X → Y where X and Y are two sets of items. It means it finds correlation between X and Y i.e. If we buy X item then it finds chances to buy Y items.

Collaborative filtering [7] faces following drawback:

a. Sparsity: the user who is very active can give rating to few items available in database. Popular item can be rate, by only a few numbers of users which is called data sparsity problem
b. Cold start: It is also known as new user problem, suppose if a user is new, then it is difficult to suggest any item,

Because no one item is rated by the user. So it is difficult to recommendation item to new user.

This paper is organized as follows: Literature survey will be presented in section 2.Problem statement is stated in section 3. Collaborative filtering is described in detail in section 4. Section 5 represents the experimental results. Section 6 summarizes the paper.

II. LITERATURE SURVEY

Today [11] the huge amount of information is available online due to the acceptance and understanding of the possibilities of internet. This [11] reason makes the World Wide Web as an important research area. Sarwar, et al., [10], has introduced and analyzed effect of different similarity algorithms and showed the experimental results through the prediction h MAE graph and also proposed that size o neighborhood effects prediction quality. Hongwu ye [7] proposed a method for finding nearest neighbor through self organizing map which makes a group of nearest neighbors which is first step in collaborative filtering. Association mining is used to fill vacant space .Thus they proposed combination of association mining and SOM to address the issue of data sparsity. Hengsong Tan, et.al; [8] presented a new approach to address the issue of data sparsity problem by combining item classification and item based collaborative approach .This approach classified the item through attributes and then produces prediction for items whose ratings are not available.

III. PROBLEM STATEMENT

A recommender system helps people who do not have sufficient personal experience to evaluate the number of alternatives offered by a Website. It provides consumer with information to help them decide which items to purchase .The proposed work is different from existing recommender systems since the existing only considers the recommending the items based on user ratings of item. It doesn't recommend items when ratings for an item are not available.

The proposed system uses combination of collaborative filtering and association mining. Collaborative filtering is used for finding similarity between items which would help the system to recommend items and association mining is used for filling the vacant ratings where necessary. Then it uses prediction of target user to the target item using item based collaborative filtering. Thus the use of both methods can help

978-1-4799-7849-6/15/$31.00 ©2015 IEEE

to manage data sparsity problem and cold start problem in recommender system.

IV. COLLABORATIVE FILTERING

Collaborative filtering is to predict [11] the users' opinion through the use of other users opinion. It uses 2 approaches:

A. Association rule prediction.
B. Recommendation.

Collaborative filtering is of two types:

1) User based Collaborative filtering: In this[10] Algorithm, similarity between, as shown in Fig.1. [11],different users are calculated by using similarity measures and then these similarity measures are used to predict ratings.It can be done through finding k nearest neighbours and then recommend items.

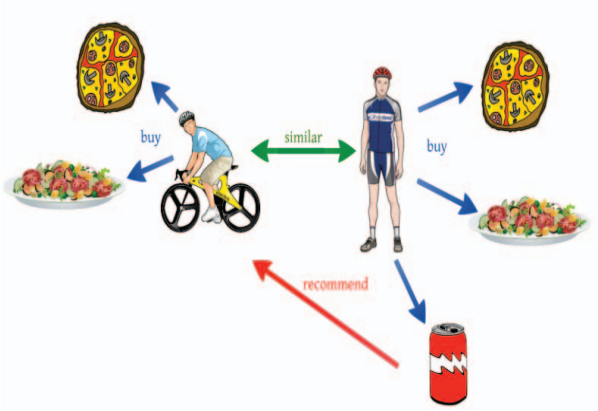

Fig.1. User based Collaborative filtering [11]

2) Item based Collaborative filtering: In this[10] Algorithm, similarity between different items,as shown in Fig.2.[11], are calculated by using similarity measures and then these similarity measures are used to predict ratings.It can be done through finding k nearest neighbours and then recommend items.

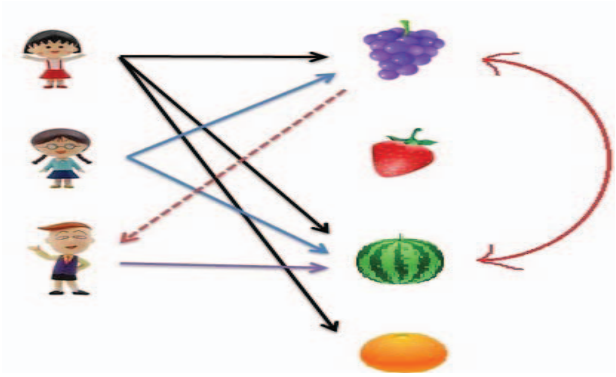

Fig.2. Item based Collaborative filtering [11]

A. How to find similarity between items.

There are many [8] similarity algorithms that can be used: Pearson correlation, Cosine Vector Similarity, adjusted cosine vector similarity.

This paper uses adjusted cosine vector similarity [8], for finding similarity among items where the difference in each users use of rating scale is taken into account

$$sim_{(t,r)} = \frac{\sum_{i=1}^{m}(R_{it}-A_i)(R_{ir}-A_i)}{\sqrt{\sum_{i=1}^{m}(R_{it}-A_i)^2(R_{ir}-A_i)^2}} \quad (1)$$

where R_{it} is the[8] rating of target item t by user i,R_{ir} is the rating of remaining item r by user I,A_i is the average rating of user i for all co-rated items ,and m is the number of all rating users to item t and item r .

Fig.3. shows the example of similarity between the items where column no 1 should ISBN no of item 1 and column 2 shows ISBN no of item2.Column 3 depicts the similarity ratio.

Item1	Item2	Sim
0195153448	0002005018	0.7214943290316204
0002005018	0195153448	0.7214943290316204
0195153448	0060973129	0.07269361923005246
0060973129	0195153448	0.07269361923005246
0195153448	0374157065	0.024886350474708552
0374157065	0195153448	0.024886350474708552
0195153448	0393045218	0.5215209699163684
0393045218	0195153448	0.5215209699163684
0195153448	0399135782	0.5943545951630751
0399135782	0195153448	0.5943545951630751
0195153448	0425176428	0.36992982183974177
0425176428	0195153448	0.36992982183974177
0195153448	0671870432	0.8379692079362827

Fig.3. Similarity Matrix Table w.r.t item – item

Now, selection of neighbors' is [3] done .Selection of neighbors is done through Threshold selection technique. This technique implies users whose similarity exceeds a certain threshold value are considered as neighbors' of target user. In our paper the threshold value we considered as 70%.

B. Prediction For Item Based Collaborative Filtering.

The rating of target user [8] u to target item t can be predicted as:

$$P_{ut} = \frac{\sum_{i=1}^{c} R_{ui} \times sim(t,i)}{\sum_{i=1}^{c} sim(t,i)} \quad (2)$$

Where R_{ui} is [8] rating of target user u to the neighbor item i, sim (t, i) is the similarity of target item t and neighbor item i and c is no. of neighbors.

V. EXPERIMENTAL RESULTS

A. *DATABASE TABLES:The tables we used in this project are:*

i. Book information table: Table I gives the following information about table.

TABLE I: Book Information Table

Attributes	**Explanation**
ISBN	Unique id of book
BOOK TITLE	Book name
AUTHOR	Name of author
YEAR OF PUBLICATION	Year of publication
PUBLISHER	Name of publisher

ii. Book Rating Table: Table II provides the information about rating given to book by user.

TABLE II: Book Rating Table

Attributes	**Explanation**
User_id	Unique id of user
ISBN	Unique id of book
Book_rating	Rating of books

iii. User information table: Table III provides the information about the user.

TABLE III: User Information table

Attributes	**Explanation**
User_id	Unique id of user
Address	Address of user
Age	Age of user

B. *Input*

Whenever we create user ,the user is assigned with user_id ,which he uses that number while logging into system . Fig.4. shows the input window

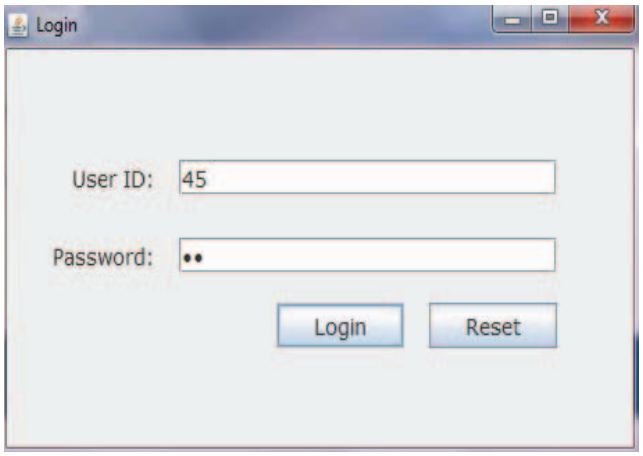

Fig.4. Input Screen

After logging into system, we need to select particular book for which we can get recommendation .Fig. 5 shows the processing window after selection.

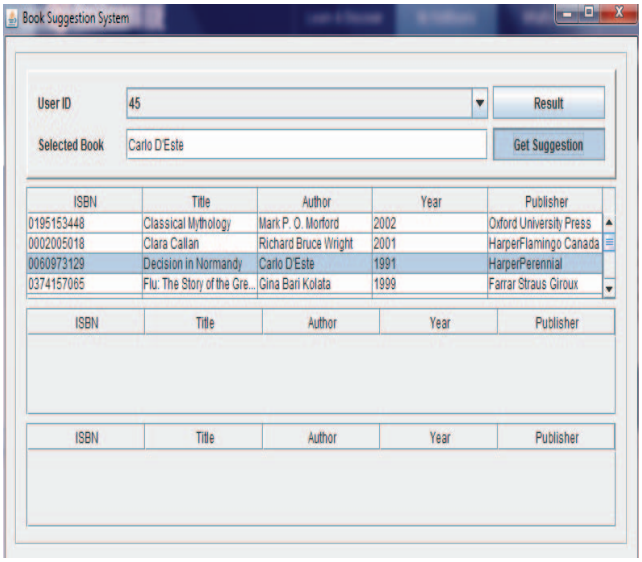

Fig.5. Processing Window

C. *Output*

Fig.6. below shows the recommendation list

a) According to Similarity calculation

b) According to author

The recommendation list which appears according to similarity calculation depends on the calculation of similarity equation which is given in eq1.The recommendation list appearing according to author depends on database.

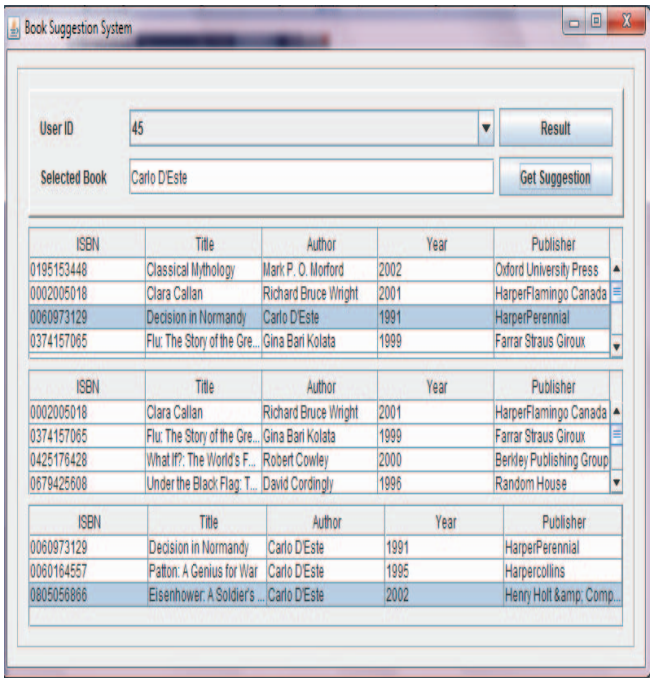

Fig.6. Output screen

D. Performance Measure :

This paper uses [8] statistical accuracy metrics where MAE (mean absolute value) is determined.

The MAE is given as:

$$MAE = \frac{\sum_{i=1}^{n} p_{i-q_i}}{n}$$

MAE is defined as average [8] absolute difference between n pairs. Assume p1….pn is predicated ratings and actual ratings are q1…qn. Lower the [8] MAE, the more accurate the prediction would be. The graph shown below verifies the accuracy of predictions in the results obtained.

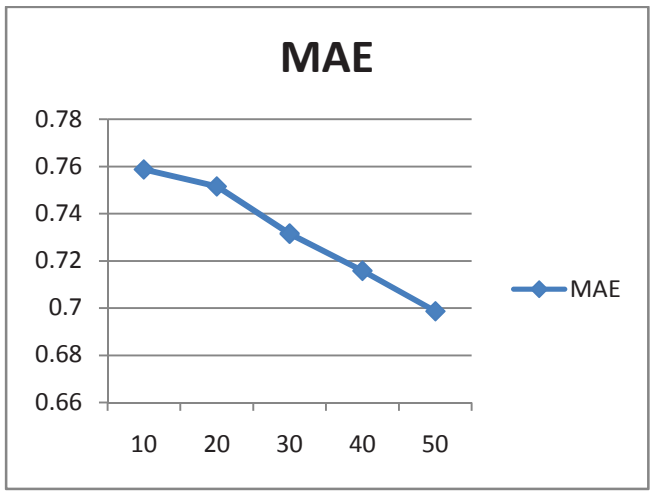

Fig.7. MAE with different no. of nearest neighbors

VI. CONCLUSION

The increasing demands of Online Information have lead to invent new techniques for prioritizing and presenting items of Users Interests. This paper uses item-based Collaborative Filtering. To produce ratings .The Item based collaborative filtering can remove the data sparsity problem and can provide good recommendation. Finally the results of similarity calculation give good performance at accuracy.

REFERENCES

[1] Zhao kai ,Lu Peng-Yu ,“ Improved Collaborative filtering approach based on User similarity Combination”,IEEE International Conference on Management Science & Engineering ,Helsinki, Finland ,2014,pp 238-241.

[2] Pijitra Jomsri, “Book recommendation System for Digital Library based on User profile by using Association rule”, Thailand, IEEE, 2014, pp 130-134.

[3] Anand Shanker Tewari, Kumari Priyanka ,”Book Re-commendation system based on Collaborative filtering and association rule mining for college students”,IEEE,2014,pp 135-138.

[4] Nirav M. Khetra, ShrutiB.Yagnik, “A Collaborative Approach for Web Personalized Recommendation System”, IJEDR, Volume 2, No. 4, 2014, pp 3761-3766.

[5] Rohini Nair, Kavita Kelkar, “Implementation Of Item And Content Based Collaborative Filtering Techniques Based On Ratings Average For Recommender Systems”, International Journal Of Computer Applications, Volume 65,No.24, March 2013, pp 1-5,.

[6] Nana Yaw Asabere, “Towards A Perspective of Hybrid Approaches and Methodologies in Recommender Systems”, International Journal of Computer Science and Telecommunications, Volume 3, No. 11, November 2012.

[7] Hongwu ye , “A Personalized Collaborative filtering Recommendation using Association rules mining and self organizing Map”, China ,Journal of Sotware,Vol 6,No 4 ,2011,pp 732-739.

[8] Hengsong Tan, Hongwu ye,” A Collaborative filtering recommendation Algorithm based on item Classfication “, IEEE Pacific Asia Conference on Circuits, Communication and system, China, 2009, pp 694-697.

[9] SongJie Gong, “Employing User Attribute and item Attribute to enhance Collaborative filtering recommendation “, China, Journal of Sotware, Vol 4, No.8, 2009, pp 883-890.

[10] Sarwar, B., Karypis, G., Konstan, J., And Reidl, J.2“Item Based Collaborative Filtering Recommendation Algorithms”, in proceedings of 10th International Conference on World Wide Web, ACM, New York, 2001, pp 285-295.

[11] www.google.com