Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 45 (2015) 23 – 31

Procedia Computer Science

International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)

Introducing Hybrid Technique for Optimization of Book Recommender System

Manisha Chandak[a], Sheetal Girase[b], Debajyoti Mukhopadhyay[c,*]

[a,b,c] *Department of IT, Maharashtra Institute of Technology, Kothrud, Pune 411038, India*

Abstract

E-Commerce has already entered into the Indian market for online shopping. People are more inclined towards online shopping which has changed the complete market scenario. There are several online shopping portals offered by organizations such as, Amazon, Flipkart, Snapdeal, Junglee, Jabong, and others, which are enjoying their online market share. As the number of online buyers and traders are increasing, effective business techniques need to be adopted to handle the large amount of data generated every day. Recommendation Systems play an important role in filtering this data and providing adequate information to the users. Various techniques like Collaborative Filtering, Content-based, and Demographic have been adopted for recommendation but there are several drawbacks causing these techniques to fail in providing effective recommendations. Therefore, it is necessary to identify more distinguishing features for optimizing these techniques. This can be achieved through utilizing the strengths of various techniques in a hybrid manner. This paper describes an effective hybrid technique for book recommendation with the use of Ontology for user profiling to increase system efficiency.

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).

Keywords: Recommender System; Hybrid Technique; Ontology

* Corresponding author. Tel.: +91-7709152655; Fax: +91-20-2544-2770.
E-mail address: debajyoti.mukhopadhyay@gmail.com

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).
doi:10.1016/j.procs.2015.03.075

1. Introduction

Recommender systems are software tools and techniques which provides suggestions to the user for the item or services useful for them. Recommender systems perform an important role in finding the customer interests[1]. This is the main reason behind their wide acceptance in most of the e-commerce businesses like online shopping and services. Today's internet user doesn't want to spend much time looking for a particular item of their interest. He/ She expect it to be taken care by the system and provide intelligent solutions. On the other hand, online traders want to know the users' interest so that they can convert these users into their long term customers. Knowing likes and dislikes of customer will always give traders an upper hand over the other traders. This gives rise to the need of recommender systems.

There are many techniques used for recommendation. These are classified on the basis of different knowledge sources. Knowledge sources include **users' features** like age and gender, **item/service features** like keywords, genre and **user-item preferences data** like rating, purchase history, etc. This user-item preference data creates a user profile which plays an important role in recommendations. Existing recommendation techniques use these user and item features for recommendation but they are not sufficient to provide effective recommendations. User profiling and improving existing techniques is the main challenge today.

2. Literature Survey

According to survey, techniques used for recommendation are classified on the basis of knowledge sources[9]. For instance, Collaborative Filtering technique works on user-item preference data, Demographic Recommendation technique uses user information while Content-based technique is based on item features.

Collaborative Filtering technique filters out the recommendation with the help of user behaviour in the form of ratings[5, 8]. This technique generates rating for an unrated item for the user, based on the commonalities among users and their ratings. Recommendation quality is directly proportional to the size of rating dataset. This technique suffers from cold start problem for a new user and a new item. This is so because user with few ratings is difficult to categorize and item with few ratings cannot easily be recommended. Also users with unusual tastes will not get recommendation as per expectations because it's quite difficult to find their co-relation with other users to extract ratings. As Collaborative Filtering recommendation works on rating history, change in user preference over a period of time affects the recommendation quality.

Demographic recommendation technique[2] uses only the information about users. This technique finds co-relation between the users based on their demographic profile. Items preferred by users having similar demographic profile are recommended. Demographic technique suffers from cold start problem for new item, as new item has not been preferred by any user of the same demographic profile. Privacy is the main concern while gathering demographic information of the users. This is due to two reasons: 1) It's difficult to get relevant recommendation when a person is looking for an item for other person belonging to different demographic profile. 2) User is reluctant to give their personal information on internet.

Content-based technique uses item features and user preference to provide recommendations[10]. In this technique, item features like keywords are used to describe items, while user preference indicates the items liked by the user. It recommends items that are similar to the items preferred by the user. This technique suffers from cold start problem for new user as user preference is not traced. Since it recommends similar items based on user preference and item feature, crossed genre items preferred by the user cannot be recommended.

Hybrid technique[3] combines two or more recommendation techniques to predict recommendation. Using Hybrid technique, it is possible to overcome the drawbacks set by one recommendation technique and sum up advantages of different recommendation techniques. For example, Collaborative Filtering technique have problem when limited user-item ratings are available whereas Demographic and Content-based technique do not use rating data and therefore can overcome cold start problem. There are various ways to combine recommendation techniques to achieve effective hybrid recommendation.

3. Methodology

The proposed system is a Book Recommender System which uses hybrid technique to predict recommendations. It

combines the features of Collaborative Filtering and Content-based technique in a mixed way i.e. it performs a union of recommendations generated by the two techniques. In addition, demographic features of users are used to filter similar users which reduce the number of comparisons used in Content-based technique to compare users.

As shown in the Fig. 1, Collaborative Filtering technique takes input from relational database to provide recommendations whereas ontology[11] is used to save user profile. This user profile information is used by Content-based technique to get input for predicting recommendations. In this recommendation system, Ontology is preferred over other databases for maintaining the user profile information gathered through browsing. Relational database systems stores data in objects like tables and views but do not store the information about how that data is related whereas, Ontology stores data in hierarchy and gives inferred results based on the data and the relation between data. It also provides simplicity in retrieving and querying data and helps in avoiding parsing of log files.

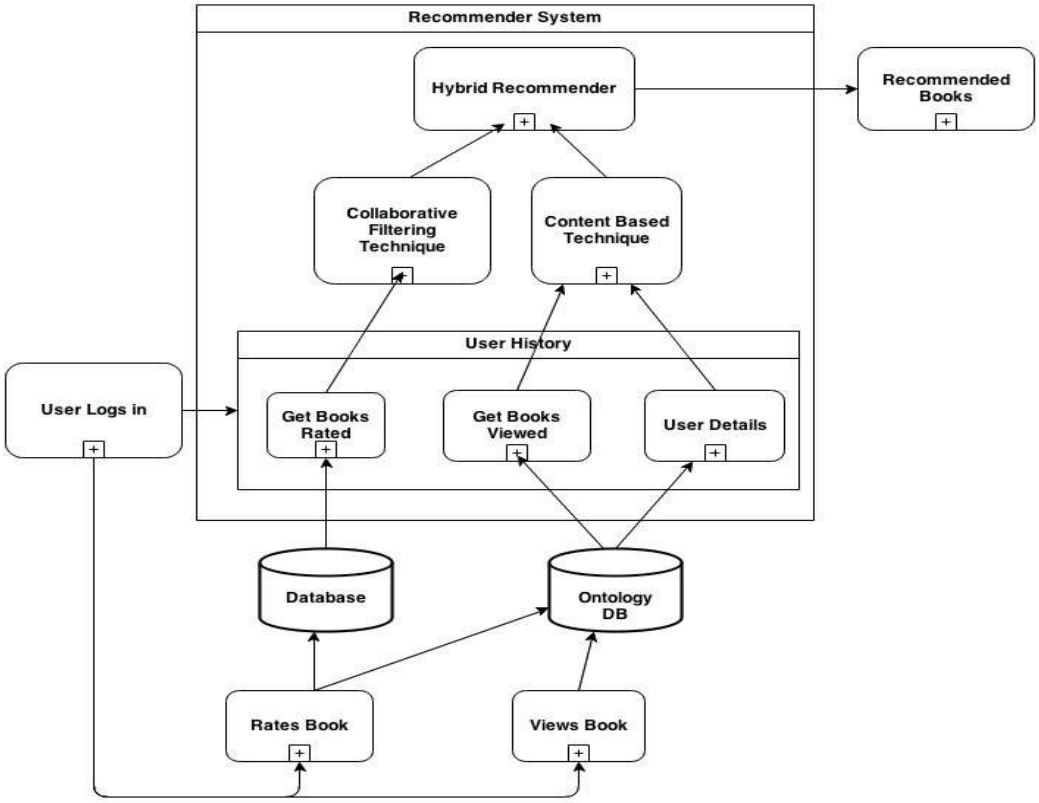

Fig. 1. Block Diagram of Recommender System

4. Book Recommendation Model

Book recommendation model uses Hybrid Recommendation technique effectively to provide recommendation to the user. In this model, hybrid technique combines Collaborative Filtering and Content-based technique along with demographic attributes. These techniques are combined using mixed hybrid concept in which several recommendations provided by different techniques are merged and presented to user at the same time.

4.1. Collaborative Filtering Technique

Collaborative Filtering technique uses user-item preferences data in the form of rating as input to provide recommendations. It predicts the ratings based on individual's past ratings and ratings given by other users stored in

a relational database. Slope one algorithm is used for prediction of ratings. This algorithm calculates the average difference between the different items and their ratings along with the number of rating for each item. If a user has rated several items, the predictions are combined using a weighted average.

4.1.1. Slope one Algorithm

Input: User, Book and Rating information from database

Output: Recommended books

Step 1: Read user book id and user rating.

Step 2: Compute frequency of rating given by users per book.

Step 3: Compute average difference in ratings given by users with respect to current user.

Step 4: Compute predicted rating as sum of the average difference in rating and rating given by current user.

Step 5: Compute weighted average rating for books with respect to frequency and predicted rating.

Step 6: Arrange the weighted average ratings in descending order.

Step 7: Display recommendations.

4.2. Content-based Technique

Content-based technique uses item features as input to provide recommendation. It recommends items by comparing them with profile of users' interest. In order to find similar item which matches the users interest MinHash technique is used which allows finding similarity among two sets. The idea is that similar items are hashed to same bucket. This states that probability of collision is higher for similar items. In this model, LSH/MinHash has been applied for finding similarity between the users. Finding similar users to the logged in user of the system and recommending books rated by those similar users is the purpose of the system.

4.2.1. LSH/MinHash Algorithm

Input: User profile (age, gender, book category) from ontology database

Output: Recommended books

Step 1: Read user profile of the current user.

Step 2: Read user profile of the other user for comparison.

Step 3: Initialize minHasharray length and i=0.

Step 4: Initialize minHash matrix to max integer value.

Step 5: Initialize minHasharray with randomly generated hash values.

Step 6: Build similarity matrix of users used for comparison.

Step 7: Calculating similarity between users

```
while similarity matrix has elements && i < minHasharray length
check
  if current user profile has element
  check
    if hashvalue< minHashmatrix[i][j]
    do
     minHashmatrix[i][j]= hashvalue
    endif
  endif
  if other user profile has element
  check
    if hashvalue< minHashmatrix[i][j]
    do
     minHashmatrix[i][j]= hashvalue
    endif
  endif
endwhile
while i< minHasharray length
check
  If minHashmatrix[0][i]== minHashmatrix[1][j]
   Identical minHash++
 endwhile
Similarity= identical minHash / minHasharray length
```

4.3. Hybrid Technique

Hybrid technique uses book recommendations given by Collaborative Filtering technique and filter out the users who rated those books. These users are again filtered out on the basis of demographic features that are age range and gender. These filtered users are compared to find similarity with current user using Content-based technique. After finding the similarity values for users, again a filter is applied which recommends books of categories viewed by the user and rated by the similar user. This reduces the number of comparison of users and makes the system effective by giving relevant recommendations. Fig. 2 explains the working of hybrid book recommendation technique.
Collaborative Filtering which looks outside the preferences of individual users provides out of the box recommendations. Content-based technique which filters recommendations on the basis of user profile along with demographic attributes overcomes the cold start problem because user preference data is not considered. Hybrid technique combines the recommendations provided by above techniques to give more effective recommendations.

5. Experiments and Results

For experiment purpose, real time data is required. Experiment uses publically available Book-crossing dataset at **http://www2.informatik.uni-freiburg.de/~cziegler/BX/** .
Since the dataset does not contain category to which the book belongs, another dataset from **https://snap.stanford.edu/data/amazon-meta.html** is used to get the relevant categories.
The Book-crossing dataset has been cleaned and transformed as:

-user(id, username, password, age, gender)
-book(book_id, book_name, book_category, book_author, book_publication, book_price, book_edition, image_url_s, image_url_m, image_url_l)
-book_ratings(book_id, user_id, rating)

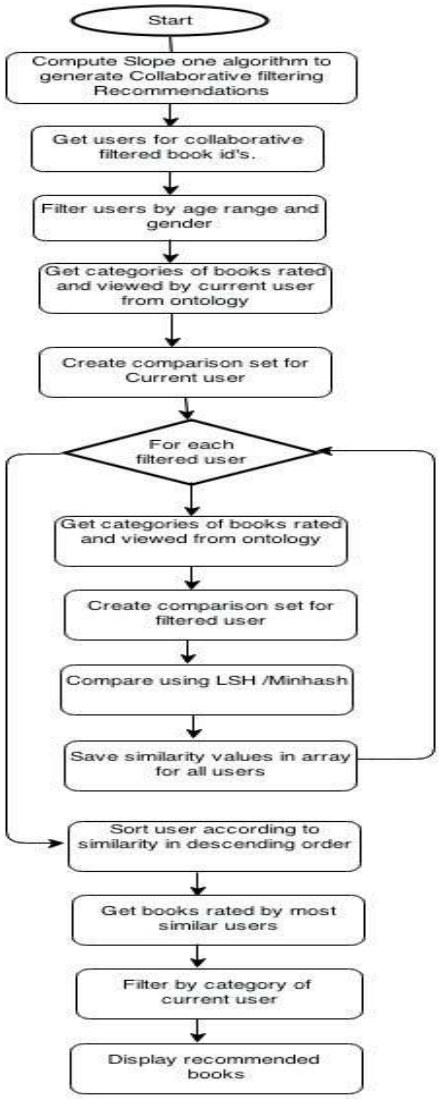

Fig. 2. Working of Hybrid Technique

Rating scale has been changed from 1-10 to 1-5. Experimental dataset now contains 6000 Users and 2600 books.

User profile has been tracked in Ontology Web Language (owl) file in this system which is used by Content-based technique for predicting recommendations.

5.1. Metrics Used For Analysis

To evaluate the recommendation techniques, three metrics have been used which are accuracy, precision and recall[6,7]. These metrics measures the system performance. In this context, **Accuracy** is the fraction of relevant recommendations to all possible recommendations, **Precision** is the fraction of relevant recommendations to the total recommendations and **Recall** is the fraction of relevant recommendations to the total relevant recommendations.

Accuracy is calculated as:

Accuracy = Relevant recommendations / Total Possible Recommendations
Precision is calculated as:
Precision = Relevant recommendations / Total Recommendations
Recall is calculated as:
Recall = Relevant recommendations / Total Relevant Recommendations
On the basis of metrics discussed above, comparitive analysis of collaborative filering, Content-based and hybrid technique have been done.

5.1.1. Discussion of Accuracy w.r.t. recommendation sytem

For experiemnt purpose, we have executed each technique separately with same data sets. As we go on increasing no. of users it has been found that Hybrid technique out performs than the other two, overcoming the drawbacks of both techniques. It was all possible due to filtering of recommendations based on demographic attributes and user preference in Hybrid technique. Fig. 3 shows the Accuracy of Recommender System for Collaborative Filtering, Content-based and Hybrid technique.

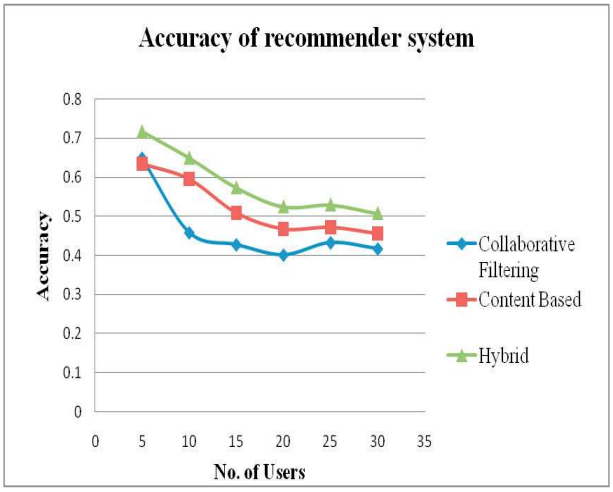

Fig. 3. Comparative analysis of recommendation techniques on the basis of Accuracy

It has been observed that:

1. Due to sparsity in ratings i.e., few users having rated the same items, the accuracy of Collaborative Filtering technique has decreased as it depends on overlap in ratings across the users.

2. Demographic attributes used to filter Content-based recommendations has add on to the accuracy of Content-based technique.

3. Sometimes Collaborative Filtering gave more relevant results due to its ability to predict out of box recommendations.

5.1.2. Discussion of Precision w.r.t. recommendation sytem

From experiemnt, it has been found that Hybrid recommendation technique has higher precision than Collaborative and Content-based technique. As we go on increasing no. of users, Hybrid technique out performs than the other two, overcoming the drawbacks of both techniques.Fig. 4 shows the Precision of Recommender System for Collaborative Filtering, Content-based and Hybrid technique.

It has been observed that:

1. As Collaborative Filtering recommends on the basis of Ratings and Content-based recommends on the basis of Item features union of both techniques increases the probability of correct recommendations. Thus, results in improved precision for Hybrid recommender technique.

2. Users with unusual taste or preference affect the Recommender System.

3. As number of users' increases, mapping of user-item ratings does not increase at same rate which results in fall in precision.

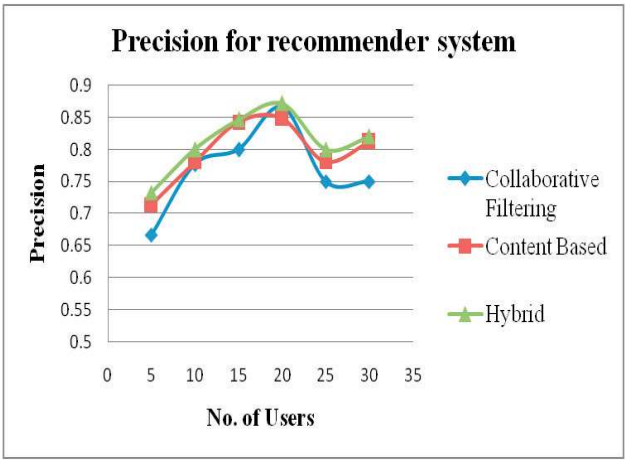

Fig. 4. Comparative analysis of recommendation techniques on the basis of Precision

5.1.3. Discussion of Recall w.r.t. recommendation sytem

From experiemnt, it has been found that Hybrid recommendation technique has higher Recall than Collaborative and Content-based technique. As we go on increasing no. of users, Hybrid technique out performs than the other two, overcoming the drawbacks of both techniques. Fig. 5 shows the Recall of Recommender System for Collaborative Filtering, Content-based and Hybrid technique.

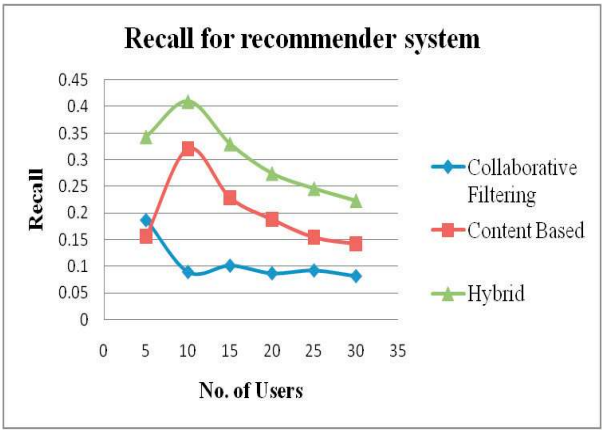

Fig. 5. Comparative analysis of recommendation techniques on the basis of Recall

It has been observed that:

1. Filtering of recommendations based on demographic attributes and user preferences, in Hybrid technique has resulted in more relevant recommendations thereby increasing the Recall of Hybrid technique.

2. Users with unusual taste or preference affect the Recommender System.

3. The number of ratings obtained is usually very small compared to the number of ratings that need to be predicted so recall decreases with increase in users

6. Conclusion

Now a day's recommendation systems are integral part of e-commerce ecosystem which demands for fast and relevant retrieval of desired information in very less time. In this scenario it is desired to have a recommendation system which reacts to huge amount of data and processing. From our experiments, we observed that combining Collaborative Filtering and Content-based technique along with demographic attributes to form a Hybrid technique gives best recommendations. Cold start problem for new users and new items has also been overcome by using demographic attributes for finding user co-relation to get recommendations. User profiling task has been simplified by making use of Ontology approach, since the overhead of parsing log files was not required

References

1. Alluhaidan, A. 2013. Recommender System Using Collaborative Filtering Algorithm, Technical Library: School of Computing and Information Systems, paper 155
2. Burke, R. 2002. Hybrid Recommender Systems: Survey and Experiments, *In: User Modeling and User-Adapted Interaction*, Volume 12, Issue 4, Kluwer Academic Publishers, pp. 331–370
3. Burke, R. 2007. Hybrid web recommender systems, In Brusilovsky, P., Kobsa, A., Nejdl, W., eds.: The Adaptive Web: Methods and Strategies of Web Personalization. *Volume 4321 of Lecture Notes in Computer Science. Springer-Verlag, Berlin Heidelberg New York*, pp. 377 – 408.
4. Gottwald, S., & Koch, T. 2011. Recommender Systems for Libraries, *ACM Recommender Systems 2011 Chicago*, IL, USA.
5. Linden, G. et al, 2003. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *In: IEEE Internet Computing*, Volume 7, Issue 1, pp. 76–80
6. Pathak, D., Matharia, S. & Murthy, C.N.S. 2012. NOVA: Hybrid Book Recommendation Engine", *IEEE International Advance Computing Conference (IACC)*, pp.977-982
7. Pathak, D. & Matharia, S. ORBIT: Hybrid Movie Recommendation Engine, IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICECCN), 2013, pp.19-24
8. Qian, F., Zhang, Y. & Duan, Z. 2013. Community-Based User Domain Model Collaborative Recommendation Algorithm, ISSN 1007-0214 03/10 pp.353-359 Volume 18, No 4.
9. Ricci, F. et al., 2011. Recommender Systems Handbook, *Springer Science + Business Media*, DOI 10.1007/978-0-387-85820-3_3, LLC.
10. Sharma, L. & Gera, A. 2013. A Survey of Recommendation System: Research Challenges, *International Journal of Engineering Trends and Technology (IJETT)* - Volume4, Issue5
11. Sieg, A., Mobasher, B. & Burke, R. 2010. Ontology-Based Collaborative Recommendation, Center for Web intelligence, Depaul University, Chicago, Illinois, USA.