ISSN (Online) 2456-1290

International Journal of Engineering Research in Mechanical and Civil Engineering (IJERMCE)
Vol 3, Issue 5, May 2018

A Hybrid Book Recommender System Using Feature Combination Technique

[1] Chitresh Kumar Singh, [2] Dr. Manash Pratim Dutta
[1][2] Dept. Of Computer Science and Engineering National Institute of Technology,Arunachal Pradesh. Naharlagun, India

***Abstract*— Recommender systems are used to provide personalized recommendations to users in the e-commerce industry. Two main approaches for the recommender systems are collaborative filtering and content based filtering. In collaborative filtering, a user's preference is calculated by his similarity to the other users. If a user has already rated or bought an item, then the preference for another user is calculated by his similarity to the other user. In content based filtering, the approach is item based, which means that if user has already rated or bought an item, then his preference for another item is based on the similarity of the first item to the second. Both of these filterings are combined in the form of hybrid recommender systems, and when weights are assigned to these recommendations, the system so developed is known as aweighted hybrid recommender system. An often neglected feature in recommender systems is that of 'Serendipity'. Serendipity means introduction of newer items into the recommender system, which are likely to interest the user. In this paper we have presented a suitable model, based on the feature combination technique, which introduces serendipity feature into the recommender systems.**

Keywords— Collaborative Recommender Systems, Content Based Recommender Systems, Hybrid Recommender, Serendipity.

I. INTRODUCTION

In recent years , recommender systems have become an important tool in the e-commerce field. They are used as a tool for gauging customer preferences and presenting options suitable to the consumer. Recommender systems are composed of three types: Collaborative, Content Based and Hybrid. In collaborative recommender systems, preferences of a user are calculated against the preferences of other users similar to the user. In case of content based systems, the user's preference for a certain item is calculated on the basis of other item that were already rated by the user. A weighted hybrid recommender system uses both collaborative and content based systems, assigning weights to the recommendations provided by both. A recommender system should also include the 'Serendipity' feature. Serendipity feature means that the user is presented with newer recommendations, rather than repetitions of similar recommendations. This provides the user with interesting new items, which results in the users' continued interest in the website/business establishment. It is necessary, as a user who is faced with repeatedly similar recommendations, will lose interest in the business. Therefore, a suitable system must be developed in order to include this feature in the recommender system. A recommender system's accuracy and efficiency may vary with different datasets. In many cases the sparsity of the dataset may cause issues in the recommender system. If we do not have enough initial ratings in order to calculate recommendations, it is known as the cold start problem[1]. It can cause inefficiency with collaborative recommender systems. In this paper, a book recommender system is being developed. The dataset used is Book-crossing dataset. The recommender system developed here is specifically for this dataset. In this research, a weighted parallel hybrid model is being used. This gives a significantly greater accuracy than the other conventional models.

II. LITERATURE SURVEY

A lot of work has been recently done in this field. Recommender systems have been developed for e-commerce companies, movies databases, jokes databases etc., each with some distinctive approach. Recent trends being developed in recommender systems are discussed by Shvartz, Lobur and Stekh[2]. The ability of users to process a large amount of given information leads to the development of software which can return relevant results as per the personalized requirements of the user. The first generation of recommender system used 3 sets, namely $U=\{u_1 , u_2 , ...u_n\}$, the set of user profile vectors, $u_t =\{x_{t1} ,x_{t2} ,....x_{tm} \}$, the set of items, $I=\{i_1 ,i_2,....i_k\}$, with item vectors as $i_v=\{i_{v1},i_{v2},....i_{vd}\}$, and the third set is $R=\{r_{ab} \}$, the set of ratings. Rows of this matrix correspond to the user set and the columns to the item set Methods

ISSN (Online) 2456-1290

International Journal of Engineering Research in Mechanical and Civil Engineering (IJERMCE)
Vol 3, Issue 5, May 2018

based on content filtering and methods of collaborative filtering are then combined in order to generate hybrid filtering methods. Various methods like Pearson's correlation coefficient, cosine correlation etc. are used to compare how similar or different two items are. Content based filtering compares items while collaborative filtering compares users. The main characteristic of these methods are that they are totally and completely dependent of statistical variables, and thereby not very accurate as such. On the other hand, G. Anatolic and L. Brkic[1] have discussed a possible solution to the cold start problem, by developing techniques to solve it using publicly available data on social media sites. Thus a recommender system may fill gaps in its knowledge and may avert or reduce the cold start problem. The authors have used Facebook API to gather up data, and further increased its value by using feedback from the system itself. Aprilianti, Mahendra and Budi[3] have implemented a weighted parallel hybrid system for e-commerce in Indonesia. They have used various sampling techniques in order to gather relevant data for the recommender system. The combination of sampling, collaborative filtering and content based filtering has resulted in a system for which the F1-measure yields a value of 9.99%. They have used the mahout framework based on hadoop to simplify their system. However, no mention has been made of the serendipity factor. It is necessary to include the serendipity factor in order to ensure customer loyalty in the process.

III. PROPOSED MODEL

The proposed model combines the features of collaborative as well as content based models. The dataset used here is the book crossing dataset, which consists of 3 tables, BX-Users, BX-Books and BX-Ratings[4]. The BX-Users corresponds to the details of the users. BX-Books contains the characteristics of the books and BX-ratings contains the ratings assigned by the users to the different books. Collaborative filtering model shall be using BX-Users and BX-Ratings, while the content based model shall use BX-Books and BX-Ratings.

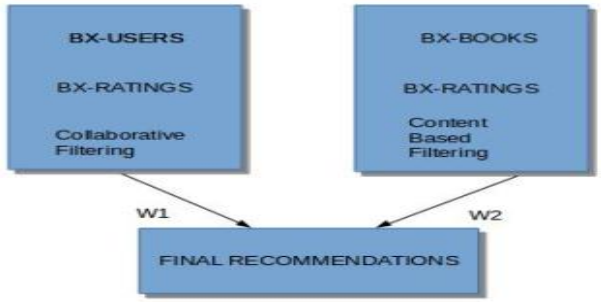

FIG. 1 DATASETS FOR FILTERING METHODS

Content based and collaborative filtering results are assigned different weights and the recommendations are finally generated. Finally, we introduce the feature of serendipity into the system. The procedure is as follows-:

- In a neighbourhood of user A if a user U gets an item X, upsetting the recommender system, increase the parameter u by a quantity h, as u=u+h, where u defines his opposition to the model.
- When U>T, a threshold value, X can be considered for random insertion.
- If a user V favours the new recommendations, he can be presented with more of such instances.
- When any user upsets the calculations by the system, the instance can be stored.
- Then by using another user's personal history, we can empirically calculate the probability of his favouring the newer recommendation.

IV. IMPLEMENTATION OF THE SYSTEM

ISSN (Online) 2456-1290

International Journal of Engineering Research in Mechanical and Civil Engineering (IJERMCE)
Vol 3, Issue 5, May 2018

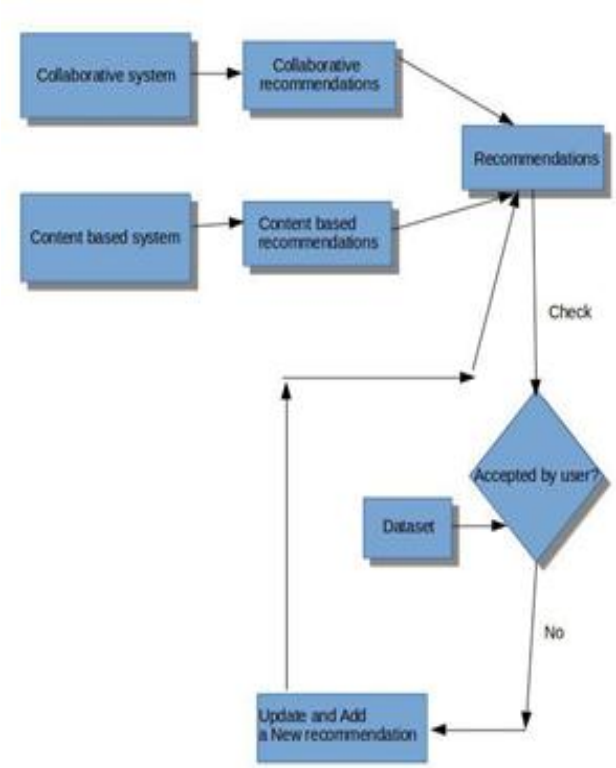

FIG. 2 FILTERING MODEL FOR THE

SYSTEM

Steps in collaborative filtering-:

●Log-likelihood similarity algorithm is used in Mahout in order to compute the similarity between target and all other users.

●Then we decide the neighbourhood threshold value to be 0.7, which is determined experimentally. Then we make a specific recommendation for this, by selecting the Top-N recommendations.

Steps in content based filtering-:

● we use the cosine similarity for this purpose.

● Then we compare different items to each other.

Lastly, we select the most accurate recommendations.

Finally, we assign weights to both the system's recommendations. Experimentally, we found that the weights 0.8 and 0.2 for collaborative and content based systems respectively yield the best combination.

Then we insert the feature of serendipity in the system.

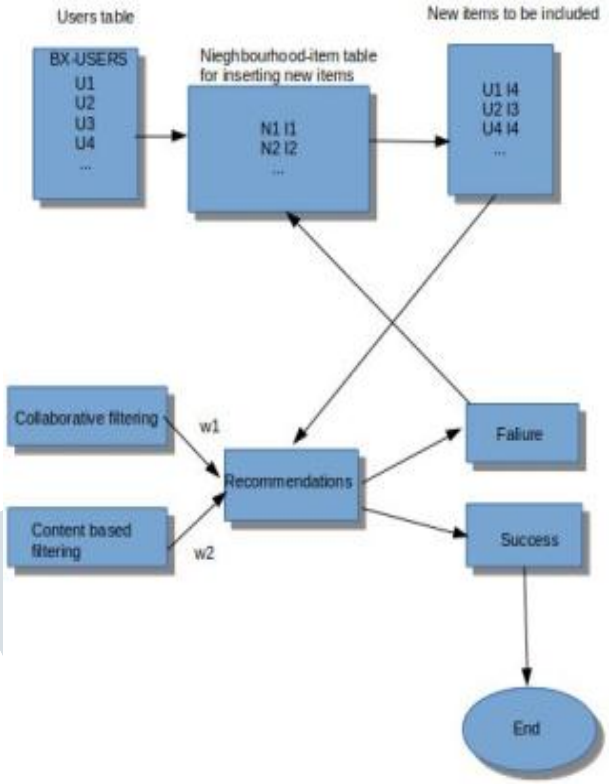

FIG. 3 INSERTING SERENDIPITY IN THE SYSTEM

V. CONCLUSION

The various results of these study are as follows:

- The assigned weights w1 and w2 for the collaborative and content based recommender systems respectively had the optimum weights as 0.8 and 0.2.
- The sampling technique used was systematic sampling.
- The optimum threshold value T is taken as 0.5.
- The F1-measure is calculated as 1.01%.

REFERENCES

[1] G. Antolic, L. Brkic, "Recommender System Based on the Analysis of Publicly Available Data", IEEExplore, MIPRO, May 2017.

[2] M.Shvartz, M. Lobour, Y. Stekh, "Some Trends in Modern Recommender Systems", MEMSTECH 2017.

[3] Mustika A., Rahmad M., Indra B., "Implementation of Weighted Parallel Hybrid Recommender Systems for e-Commerce in Indonesia", IEEExplore, ICACSIS 2016.

[4] Cai-Nicolas Ziegler, "Book-Crossing Dataset", http://www2.informatik.unifreiburg.de/~cziegler/BX/,August /September 2004.

[5] Toby Segaran, "Programming Collective Intelligence", J. O'Reilly Media, Inc., Aug.2007, ISBN-10:0-596-52932-5.