Recommender System using Hybrid Approach

Sanya Sharma[1]	Aakriti Sharma	Yamini Sharma	Ms. Manjot Bhatia
Student, Department of IT	Student, Department of IT	Student, Department of IT	Associate professor, Department of IT
JIMS, Delhi	JIMS, Delhi	JIMS, Delhi	JIMS, Delhi
[1]s.sanya1807@gmail.com	aakritisharma1992@gmail.com	yamini.sharma.1910@gmail.com	Manjot.bhatia@jimsindia.org

***Abstract* - Efforts from multiple disciplines involving various fields like Human Computer Interaction, Marketing, or Consumer Behavior, Data Mining, Adaptive User Interfaces, Artificial intelligence, Statistics, Information Technology and Decision Support Systems contribute to development of recommender system. In today's scenario, there exists different recommender algorithms used for filtering data and providing user with best suitable choices. The way people find products and information is greatly affected by recommender systems. Recommender system is thus, a tool to reduce the overloaded data. In this paper, we propose a new algorithm Composite Search that combines few filtering algorithms and presents refined result, eliminating drawbacks of other algorithms. We present our approach that processes data and provides more filtered result.**

Keywords-Recommender system, content-based, collaborative, hybrid, nearest neighbors, filtering.

I. Introduction

In the era of technology, data filtering techniques have altered the issues to larger extent in e-commerce and social media. Recommender system, a tool used for filtering data, modified the manner one will notice information, merchandise and other forms of data from large quantity of data repository. It uses opinions of users of explicit communities to assist people therein community to spot content of interest effectively from a set of selections [2]. It has modified the manner people share their information, opinions and data with different users. This is applied to varied social parts (e.g. events, folks or groups) sharing information or item based mostly (TV program/show/episode, movies, images, music, news, web pages, scientific literatures, books etc.) that are of user's interest. User's profile (with relevant personal information as well as search history of a user) is compared with search history of different users for a few attributes of the item user is looking for and on the basis of 'ratings' or 'preferences' given by different users for that item, recommender system predicts the recommendations for that item. The filtering techniques
have been categorized as 1) Content-based, 2) Collaborative and 3) Hybrid. These are discussed below:

A. The Content-based Approach

Content-based filtering technique works with profiles of users. A profile has information about a user and his preferences. Preferences rely on how the user rated items and what item user has bought and viewed. Generally, once profile is made, recommender systems create a survey, to get basic information about a user so as to avoid the new-user problem. New-user problem arises when profile of the same user is made having different attributes. Profiles are obtained by analyzing the items previously seen and rated by the user and are typically created using keyword analysis techniques from information retrieval. Within the recommendation method, the engine compares the things that were already positively rated by the user with the things that didn't rate and appears for similarities. Those things that are mostly the same as the positively rated ones will be suggested to the user. [1]. Fig 1 shows an example of a user profile with the mobiles user has already viewed and the ratings the user has made. Fig 2 shows the list of mobiles with its attribute-values. A content based recommender system would find out mobile from the list (Fig 2) that the user has already viewed and rated positively. Then, it will compare those mobiles with the rest of the mobiles from the list (Fig 2) and look for similarities. Similar mobiles will be recommended to the user. In the current example we can see that there is a mobile "Samsung Galaxy A5" similar to the mobile" Samsung Note 4" that had been positively rated by some other user. The user hasn't rated "Samsung Galaxy A7" so it will be suggested to him.

ISBN: 978-1-5090-1666-2/16/$31.00 ©2016 IEEE

Mobiles	*Lumia 535*	*Samsung GalaxyA7*	*Samsung Note 4*	*iphone 6*
Ratings	6	7	8	9

Fig 1: The phones the user has viewed

Mobiles	*OS*	*Rear camera*	*Battery*	*2G/3G/4G*	*Touch*
Lumia 535	7	7	8	8	7
Samsung Galaxy A5	6	5	7	8	7
Samsung Note4	9	9	8	9	10
iphone6	10	10	10	10	10

Fig 2: List of mobile phones

There are vivid algorithms for measuring similarities among items in database and those in user's profile
[4]. One of such approaches is cosine similarity. The measured angle between items, where items are represented as vectors gives out their cosine value as shown in Equation 1. Vectors of two items with attributes are compared in cosine similarity function as follows (Fig 3) [6].

$$u(c,s) = \cos(\overrightarrow{w_c}, \overrightarrow{w_s}) = \frac{\overrightarrow{w_c} \cdot \overrightarrow{w_s}}{\|\overrightarrow{w_c}\| \times \|\overrightarrow{w_s}\|}$$

$$= \frac{\sum_{i=1}^{K} \vec{w}_{ic}\vec{w}_{is}}{\sqrt{\sum_{i=1}^{K} \vec{w}_{ic}^2}\sqrt{\sum_{i=1}^{K} w_{is}^2}} \qquad (1)$$

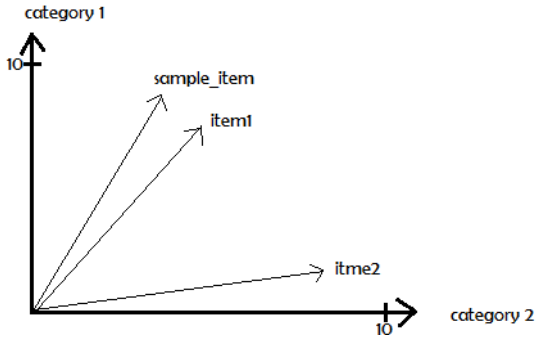


Fig 3: cosine similarity on cordiante plane

The more similar two items are, smaller is the angle between their vectors and for less similar items, the angle will be large [7].

B. The Collaborative Approach

In collaborative technique, the concept of neighborhood is used. Here, neighbors are the group of people who have rated same item and have similar taste. Items are recommended which have not been rated by the user but were already positively rated by users in his neighborhood. In Fig 3, Samsung Galaxy A7 and Samsung Note 4 are positively rated and have similar attributes therefore they make a neighborhood. In Fig 4, the mobile "Samsung Note 4" was positively rated but has not been viewed by the user A so it will be recommended to user A.

User/Mobiles	*Lumia 535*	*Samsung Galaxy 5*	*Samsung Galaxy A7*	*Moto E gen2*	*Iphone 6*	*Samsung Note4*
User A	6	7	8	6	10	-
User B	7	8	7	8	9	9
User C	8	8	9	8	10	10

Fig 4: Example of collaborative approach

Collaborative filtering is not only used in e-commerce but also in browsing of certain documents (e.g. documents among scientific works, articles, and magazines).

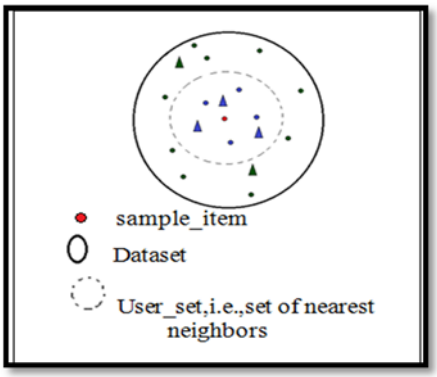


Fig 5: Nearest Neighbor Algorithm(Collabrative Filtering)

We can distinguish the collaborative filtering on basis of following famous approaches: i) user-based ii) item-based.

i. User-based approach

The main role is performed by user in the user-based approach. Items will be recommended to the user based on what is viewed by other user from the same group. Recommendations are given to the user based on evaluation of items by other users from the same community, with whom user shares common preferences. If the community rates the item positively, it will be suggested to the user. Thus in the user-based approach the items that were already rated by the user plays an important role in searching a group that shares preferences with him [9] [10] [11].

ii. Item-based approach

It is fact that the preferences of users remain constant or change very slightly so the similar items build neighborhoods based on appreciations of users. Afterwards the system generates recommendations of items in the neighborhood that a user will prefer [12] [8].

iii. The Hybrid Approach

This approach combines the multiple filtering methods to get a refined result. Different criteria to combine different recommendation filtering techniques are classified as: [12]

1. Separately implementing content-based methods and merging their predictions,
2. Integrate some content-based characteristics into a collaborative approach,
3. Integrate some collaborative characteristics into a content-based approach, and
4. Developing a general integrated model that merges characteristics of both content-based and collaborative filtering methods.

It reduces the filtering problems as advantages of one technique can be used to minimize the disadvantage of other technique.

II. Literature Survey

As mentioned and described by Paul Resnick and Hal Varian in 1997, the collaborative filtering approach is one of the most researched techniques of recommender system. In collaborative filtering technique, the community of people is formed who have common preferences in their profiles and the items positively rated will be recommended to the people within that community. There arises dynamic problem with the recommender system where user does not want his personal choices, reviews or interests to be share publically. In general, the recommendations can be better analyzed when more information is provided regarding recommendations. This problem can be overcome with providing and sharing information anonymously but it cannot solve the problem completely. There are various business models for recommender systems that would help generating revenue as the maintenance for the system is cost effective [2]. Sparsity of the user-item rating dataset is a major problem in collaborative filtering which leads to poor traditional approach of comparison of data items. Solution to this problem was proposed in which user rating can be pre-produced on the bases of item classification. Items are classified to predict the user ratings of the spared values and then recommendations are produced by applying item-based collaborative filtering. It proved to be the better solution to the traditional collaborative filtering [4].The K-nearest neighbor (KNN) method makes collaborative filtering more realistic. In this approach for the given user its k-nearest neighbor will be considered which likely to have more common interests. In user-based clustering algorithms, cluster of similar items is formed from which top-N neighbors are chosen. This approach has low accuracy of prediction but high scalability [3]. To remove these drawbacks user partition approach was introduced which provides better accuracy and performance. Users are partitioned on the basis of their preferences and using those preferences item is predicted [13].Daniel Lemire in 1995 proposed SLOPE ONE algorithm which is based on new item based collaborative filtering. It is more accurate for sparse data [10]. Cloud computing has overcome the problem of large scale computation. It provides virtualization of resources over internet. Implementation of collaborative filtering technique on cloud computing has solved the problem of scalability. One of the cloud computing platforms is Hadoop which uses distributed file system. Using hadoop one can easily run programs in parallel and it improves speed by dividing big problem into small segment which can be handled by hadoop. The user-based collaborative filtering for reduced map is implemented on hadoop platform. This implementation is tested for various configurations and it is found that for simple program linear speedup can be achieved [10].

III. Proposed Composite Search Algorithm

The Proposed Composite Search Algorithm is based on hybrid approach, overcomes the disadvantages of one filtering technique by applying the other filtering technique on the result of the former technique. For instance, the item which has not been rated cannot be recommended in collaborative filtering. In content-based filtering, since the prediction is based on the attributes or the description of the item, there will be limit to the recommendations. Therefore two or more techniques can be merged in such a way which can recommend as much possible similar items.

A. Algorithm

Let us consider an algorithm where i=0, 1, 2, 3, .n denotes items of actual dataset, sample_item is the item which user wanted to view or buy. Distance between the two items can be measured on the basis of similar categories that two items share. j=0, 1, 2, 3, .n denotes the items of subset of actual dataset which consists of items that are nearest neighbors i.e., they share similar categories.

COMPOSITE SEARCH ALGORITHM

Input: Item to be searched.

Output: Items which are similar in attributes with the input item as well as positively rated by other users.

STEPS	PROCESSING
Step1:	Visit each i^{th} item of dataset and calculate distance (similar attributes) from i^{th} item of dataset to sample_item.
Step2:	i^{th} item having least distance (nearest neighbors) from sample_item, mark them as user_set (see Figure5).
Step3:	Apply cosine similarity function on every j^{th} item in user_set and sample_item(see Figure3)
Sep4:	Classify every j^{th} item of user_set (subset of dataset) having value greater than zero of cosine function and positive rating or positive user feedback (i.e., rating>=3) given by other users as recommendation set.

Fig 6: Composite search algorithm

B. Privacy concern and security issues

Recommender systems involve an inherent trade-off between accuracy of recommendations and the extent to which users are willing to release information about their preferences. Using online applications users may share or upload their personal information but this information is shared within the specific scope. The privacy of the information means exposure of the information within a bounded scope. This scope is defined the size of the audience, duration and the extent of usage allowed define the scope of information. Unauthorized and unwanted access to the user information is privacy breach. Personalized recommender systems take user preferences as input in order to make recommendations. Providing more preferences improves the accuracy of the recommendations but it may also increase the risk of privacy breach. More preferred recommender is one which takes minimum number of inputs and gives as possible as recommendations. Users' profile are the main source from

the preferences of the user can be obtained .Since the users' profile and the profiles of users' neighbors are scanned every time an input is given. This leads unwanted exposure and access to user's own personal information [14].

In composite search, user is not required to give up all his/her information to the recommender. When an input is provided, the system matches the ratings and minimum three attributes (i.e., which is considered as the positive rating) with the other items which are being purchased, rated, viewed etc. by neighbors of the user who has provided the input. As soon these constraints have been fulfilled the system may stop scanning further in user's profile. To some extent it may reduce the exposure of users' personal information. Hence preserves the privacy.

IV. Analysis

Different websites use different data filtering techniques. Recommendation results of some of the websites have been compared with results generated by composite search algorithm. In the following analysis, the user input has been taken, each word mentioned in the user input is treated as attribute be it type, color, brand, size etc. On the bases of attributes and the rating given to the items by other user the recommendations are generated. The comparison is given below:

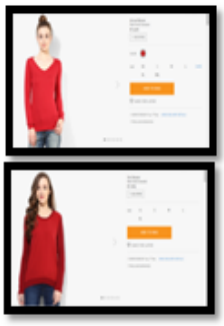
	We searched for the item "**Vero Moda Red Sweater**"
	**Based on the algorithm used by Jabong.com**, we got the following recommendations on the following parameters: • Brand Name • Sweater • No color specifications in recommendations
Recommendations by composite search algorithm : 	**Now on applying composite search algorithm**: 1. First we apply nearest neighbor 2. Applying the cosine similarity, let the condition be "**color**" 3. Now filter on user rating i.e. the user must have provided a rating to the item and moreover it should be greater than 3

Fig 7: Jabong analysis

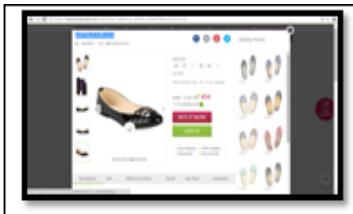
	We searched for the item "**solid black ballerinas**"
 	**Based on the algorithm used by Limeroad.com**, we got the recommendations on different category which includes: • Similar Picks • Type of shoe • Similar Price No color specifications in recommendations
Recommendations by composite search algorithm :   	**Now on applying composite search algorithm**: • First we apply nearest neighbor • Applying the cosine similarity, let the condition be "**Ballerinas**" • Now filter on user rating i.e. the user must have provided a rating to the item and moreover it should be greater than 3

Fig8: Limeroad analysis

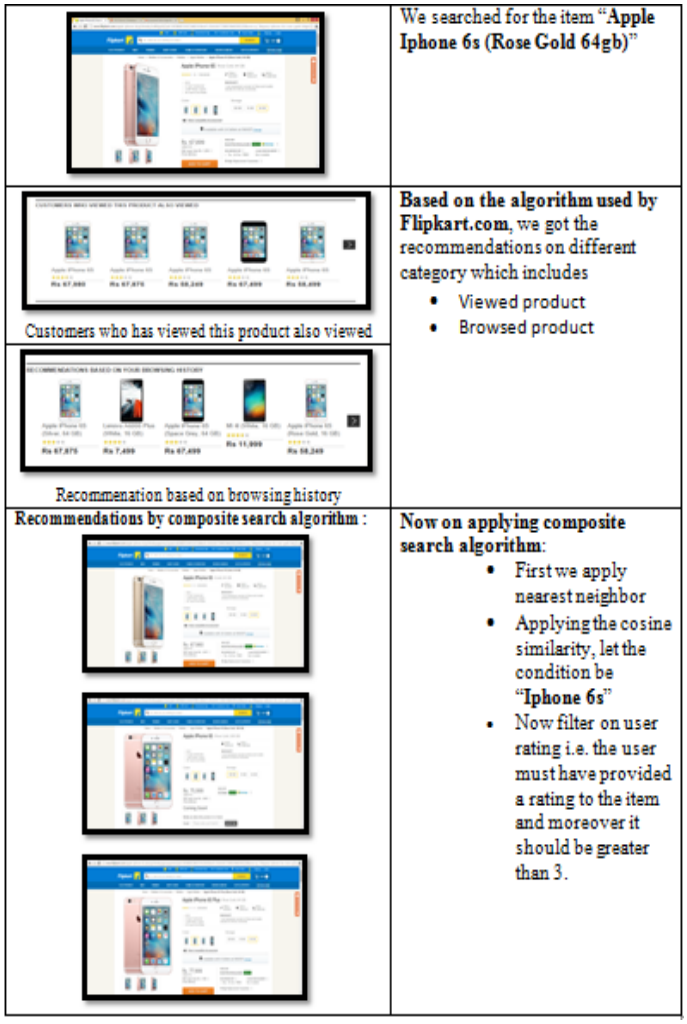


Fig 9: Flipkart analysis

V. Conclusion and Future Work

Recommender system opens many new options for searching and filtering information. It is of ever increasing importance to the sphere of e-commerce. Indeed, a good recommender system is equivalent to a strong marketing or advertising campaign. It has the potential to multiply sales numbers and engender brand loyalty in the customer base who grow to trust and rely on the recommendations put forward by the system. This paper proposes composite search algorithm that is based on:

(i) Cosine similarity function (ii) Rating given by other users. According to analysis, recommender algorithms of other websites work on the bases of either attributes or ratings given by other users. Proposed algorithm refines data on the bases of attributes as well as user given ratings. This provides improved recommendations to the users. History of user plays an important role in recognizing user's interests. Composite Search does not take user's search history into consideration. So we are planning to consider this feature in our future work.

References

[1] Daniar Asanov, Algorithms and Methods in Recommender Systems, Berlin Institute of Technology, Berlin, Germany, 2011.

[2] P. Resnick and H. R. Varian, "Recommender systems," Common. ACM, vol. 40, March 1997, pp.56–58.

[3] Jianyong Wang, Jiawei Han, Ying Lu and Petre Tzvetkov "TFP: An Efficient Algorithm for Mining Top-K Frequent Closed Item sets", IEEE transactions on Knowledge and Data Engineering, Vol.17, No.5, May 2005 pp.652-664.

[4] H. Tan and H. Ye, "A collaborative filtering recommendation algorithm based on item classification," in Proceedings of the 2009.

[5] Kuepper, "Recommender systems," ecommerce, vol. Chapter 08, p. 13, 2011.

[6] A scalable collaborative filtering algorithm based on localized preference, vol. 1, 2008.

[7] Z. Sun and N. Luo, "A new user-based collaborative filtering Algorithm combining data-distribution," in Proceedings of the 2010.International Conference of Information Science and Management Engineering - Volume 02, ser. ISME '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 19–23.

[8] S. Vucetic and Z. Obradovic, "A regression-based approach for scaling up personalized recommender systems in e-commerce," in In: ACM WebKDD 2000 Web, 2000.

[9] Z.-D. Zhao and M.-s. Shang, "User-based collaborative-filtering Recommendation algorithms on hadoop," in Proceedings of the 2010 Third International Conference on Knowledge Discovery And Data Mining, ser. WKDD '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 478–48.

[10] X. Zhu, H.-W. Ye, and S. Gong, "A personalized recommendation System combining case-based reasoning and user-based collaborative filtering," In Control and Decision Conference, 2009. CCDC '09. Chinese, 2009, pp. 4026 –4028.

[11] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based Collaborative filtering recommendation algorithms," in Proceedings of The 10th international conference on World Wide Web, ser. WWW '01. New York, NY, USA: ACM, 2001, pp. 285–295.

[12] Lalita Sharma, Anju Gera, and A Survey of Recommendation System: Research Challenges International Journal of Engineering Trends and Technology (IJETT) - Volume4Issue5- May 2013 ISSN: 2231-5381, Page 1989.

[13] Liang Zhang, A scalable collaborative filtering algorithm based on localized preference, Machine learning and Cybernetics vol. 1, July2008.

[14] Shyong K. "Tony" Lam, Dan Frankowsk, John Riedl, Do You Trust Your Recommendations?, An Exploration of Security and Privacy Issues in Recommender Systems, International Conference, ETRICS 2006, Freiburg, Germany, pp 14-29, June 6-9, 2006.