International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 4, Issue 3, March 2015 **ISSN 2319 - 4847**

Survey on Collaborative Filtering Technique in Recommendation System

Omkar S. Revankar, Dr.Mrs. Y.V.Haribhakta

Department of Computer Engineering , College of Engineering Pune, Pune, India

ABSTRACT

This article presents an overview of recommendation System and illustrates the present generation of recommendation techniques. These techniques are usually categorized into three main classes as Collaborative Filtering (CF), Content-Based Filtering (CBF), and Hybrid Recommendation approaches. CF is a framework for filtering information based on the preferences of users. This technique can predict a user's preferred items by using the user's known history data as well as other users' known history data, and then recommends items to the user. This paper is focused on collaborative filtering; its types and its major challenges for instance cold start problem, data sparsity, scalability and accuracy.

Keywords:- Recommender systems, Collaborative Filtering, Content based Filtering.

1. INTRODUCTION

Due to tremendous increase in e-commerce and online web services the matter of information search and selection has become increasingly serious; users are confused by which alternative to consider and they may be not having sufficient time or knowledge for personal evaluation of these alternatives. Recommender systems have proved that , they are a helpful way for online users to deal with the information load and have turn out to be one of the most popular and powerful tools for e-commerce. Recommender system provide list of items by predicting which item are most suitable to user, based on user past history, preferences and constraints [4]. Different algorithm and methods are used by recommendation systems to provide personalized recommendations.

Recommendation systems are divided into following three categories:-

Content-based Recommending

In this method user will be recommended items comparable to the ones the preferred in the past, or matched to the user characteristics.

Collaborative Filtering

It is the process of evaluating or filtering items using the opinions of other users. Collaborative filtering techniques collect user's profiles, and examine the connection among the data according to similarity function. The likely categories of the data in the profiles include user behavior patterns, user preferences, or item properties. Collaborative filtering technique collects large information about user behavior, history, click pattern and recommends what user will like based on
his similarity with other users communally.

Hybrid methods

Hybrid methods are a combination of different filtering methods. In Hybrid collaborative filtering [5], [6] system, collaborative filtering is combined with other recommendations techniques like content based filtering. Content based recommendation system make predictions based on the content of textual information like URLs, logs, item description and profiles about user taste, preferences and needs. Demographic based recommendation system [8] makes use of user profile information such as occupation, gender and postcode. Utility based recommenders and knowledge based recommender system uses knowledge about how a particular object satisfies user needs.

2. COLLABORATIVE FILTERING TECHNIQUES

Collaborative filtering is widely used in recommender systems. Collaborative filtering(CF) attempt to mechanize "word-of-mouth" recommendation procedure that means, the objects are suggested to the customer according to how customers with similar interests, categorize these objects.[7] Collaborative filtering technique collects large information about user behavior, history, click pattern and recommends what user will like based on his similarity with other users. For example, Amazon's recommendation algorithm collects items which are similar to purchases of a user and ratings, without ever calculating a predicted rating [3].

Collaborative filtering techniques can be categorized into two types:

(a) Memory Based Collaborative Filtering Technique

Memory based collaborative filtering techniques use item-to-item or user-to-user correlations to make prediction for user on future items. For computing prediction, whole training set is taken into memory, making it easier to include new data but experiences slow performance on large information datasets. This issue can be overcome by pre-calculating correlations and updating it. Memory based collaborating filtering technique are categorized into two types according to "Nearest neighbor algorithm" [15].

(1) Item based filtering technique

It mostly focuses on most similar items. The key idea is that users are likely to have same opinion for similar items [1]. Similarity between items is decided by looking at how other users have rated items. Item based filtering technique overcome the problem of user cold-start problem and partially improves scalability problem as similarity between items is more stable than between users.

- Item Based Algorithm

for each item k for which user h has done no choice yet

for each item m for which user h has done choice so far

- calculate similarity c (using Pearson correlation/Cosine similarity) between k and m
- -add h's choice for m, according to ranks generated by similarity c.

return the top N ranked items

A prediction for a user u and item i is composed of a weighted sum of the user u's ratings for items most similar to i where itemSim()is a measure of item .ItemSim() is calculated using formula (1) below:

$$itemSim(i,j) = \frac{\sum_{uCRBi,j}\left(r_{ui}-\bar{r}_u\right)\left(r_{ni}-\bar{r}_n\right)}{\sqrt{\sum_{u\subset RB_{i,j}}\left(r_{ui}-\bar{r}_u\right)^2}\ \sqrt{\sum_{u\subset RB_{i,j}}\left(r_{nj}-\bar{r}_u\right)^2}} \tag{1}$$

Adjusted-cosine similarity, similarity metric is used to compute item similarity by using all users who have rated both item i and j as in formula (2) below:

$$pred(u,i) = \frac{\sum_{j \in ratedItems(u)} itemSim(u,n).(i-\bar{j}).r_{ui}}{\sum_{j \in ratedItems(u)} itemSim(i,j)} \tag{2}$$

- Problems of item based collaborative filtering

(i)Item Cold-Start problem– Cannot predict ratings for new item till some similar users have rated it, this problem also occurs in user based collaborative filtering technique but it is a bigger problem here.

(ii) As the number of items increases the size of the model increases. We can reduce the size by only storing correlations for item pairs with more than k co- ratings, by pruning many of the correlations it becomes difficult to make a prediction for a given target item and user, since the items correlated with the user's ratings may not contain the target item.

(2) User based filtering technique

It mostly focuses on most similar users. Recommendation system based on user based filtering technique generates prediction based on ratings from similar users called as neighbors.

- User Based Algorithm

for each item k for which user u has done no choice yet

for each other user v who has done choice for item k

- calculate similarity c (using Pearson correlation/Cosine similarity) between u and v
- add v's choice for k, according to ranks generated by similarity c .

return the top N ranked items.

If a user n is having similarity with a user u, we say that n is a neighbor of u. User-based algorithms generate a prediction for an item i by analyzing ratings for i from users in u's neighborhood by using the following formula (3).

$$userSim(u,n) = \frac{\sum_{iCRu,n}(r_{ui}-\bar{r}_u)(r_{ni}-\bar{r}_n)}{\sqrt{\sum_{i\subset CR_{u,n}}(r_{ui}-\bar{r}_u)^2 \sum_{i\subset CR_{u,n}}(r_{ni}-\bar{r}_n)^2}} \quad (3)$$

userSim()is calculated by using the Pearson correlation. The Pearson correlation coefficient is calculated by comparing ratings for all items rated by both the target user and the neighbor as given in following formula (4). Pearson correlation ranges from 1.0 for users with perfect agreement to -1.0 for perfect disagreement users [1] [16].

$$pred(u,i) = \bar{r}_u + \frac{\sum_{n\subset neighbors(u)} userSim(u,n).(r_{ni}-\bar{r}_n)}{\sum_{n\subset neighbors(u)} userSim(u,n)} \quad (4)$$

- Problems Of User-Based Algorithms

(1) User Cold-Start problem- not enough known about new user to decide who is similar.

(2) Sparsity when recommending from a large item set, users will have rated only some of the items which makes it hard to find similar users.

3] Scalability– with millions of ratings, computations become slow.

4] Item Cold-Start problem– Cannot predict ratings for new item till some similar users have rated it.

(b) Model Based Collaborative Filtering Technique

User recommendations in model-based collaborative filtering algorithms are based on learned models. Model-based collaborative filtering analyze the training data, summarize the complicated patterns into the learned models, and then make predications based on the learned models. Results from model-based CF usually have less accuracy in prediction as compared to memory-based methods on dense data sets where user-item values are available in a large fraction in the training set, but give better performance on sparse data sets.

Model based collaborating filtering techniques are categorized into two types:

(1) Cluster Model

Multinomial mixture model is also called as cluster model in which the probability is conditionally liberated from participation votes in a class C variable obtaining several relatively little numbers of values that are discrete. In cluster model, collection of preferences is taken by definite groups or kind of users that are similar in their clusters.

In the certain classes, the preferences that are related to the dissimilar items are liberated. The cluster model defines the probability of joint probability of votes and class to a group of marginal and conditional distribution [18].

Pr (C=C1V1,C2V2,…,Vn) = Pr (C=C)Pr (Vi |C = C)

Pr (C=C1V1,C2V2,…,Vn) = Probability of a person of specific class and a collection of votes.

Pr(C=C) = Class membership probability.

Pr (Vi |C =C) = From user's vote training data condition probability of votes is calculated.

If no variable is seen in user database then we have to use methods that can learn parameters with hidden variables for modeling.

(2) Bayesian Network Model

Bayesian network model is represented in a directed acyclic graph. This model is represented in a triplet<M.D,@>, where M denotes random variable, D denotes directed arc and @ denotes conditional probability table determining how much a node depends on its parent node[19].

***(*1) Simple naïve Bayesian collaborative filtering**

In Naive Bayes classifier we assume the features are independent for a given class, the probability is calculated by taking all features, the class with the highest probability will be classified as predicted class.

(2) TAN-ELR collaborative filtering

Simple naïve Bayesian collaborative filtering method lacks when dealing with partial data or in completed data. This technique is overcome by Extended Logistic Regression algorithm(ELR) which is a discriminative parameter learning algorithm that maximize log conditionals likelihood TAN-ELR (Tree Augmented Naïve Bayes) have high classification accuracy for both complete and incomplete data[17].

3. CHALLENGES OF COLLABORATIVE FILTERING

Scalability

In any of the environments where collaborative filtering is used in recommendation, [2] there are millions of users and items this creates problems as there are inadequate computational resources to meet the new demands of user. Large Computational Power is required to produce recommendations. For an example if there are M users(Millions of user's) and N(Millions of item's) items then overall complexity of an algorithm will be Big-O(n)these is very large complexity with that many systems have to provide quick online recommendation for all user's regardless of their purchase and rating history.

Data Sparsity

In E-commerce system with large number of products, size of user-item matrix becomes large and sparse. Due to this it becomes difficult to make recommendations and maintain the quality of recommendation system. The data sparsity problem is also known as "Cold Start" problem. As explained in [7] [13], the cold start problem arises in a situation where a new user or item has just registered into the system and it becomes difficult to find neighbor users or items as inadequate information about user and item is available, and new users are unlikely given good recommendation due to lack of ratings and purchase history.

Synonym problem

Some recommendation system categories item into different category because of difference in names.

Gray sheep

Some users cannot take benefit of collaborative filtering technique as they will not consistently agree or disagree with group of people, [2] these people are called Gray sheep people.

Black sheep

Blacksheep are opposite group of people who are not in favor of recommendation system concept.

4. RELATED SURVEY

A review of our study on following research work shows that ,the authors have used various collaborative techniques like personalized collaborative filtering, user based collaborative filtering, item based collaborative filtering etc. for building a recommender system. Some drawbacks of collaborative filtering such as cold-start problem, data sparsity and scalability are also discovered and overcome. Yechun Jiang, Jianxun Liu, Mingdong Tang and Yechun Jiang, Jianxun Liu, Mingdong Tang (2011) [9] discussed on personalized collaborative filtering method used for Web service recommendation. Computation of similarity is a main part of Web service recommendation techniques. Similarity measurement is calculated among users and personalized influence of services. By utilizing the similarity measurement model of Web services, an effective Personalized Hybrid Collaborative Filtering technique is developed by integration of personalized user-based algorithm and personalized item-based algorithm. They also conducted experiments based on real Web service QoS dataset WSRec which consists of more than 1.5 million test results of 150 service users belonging to different countries on 100 publicly available Web services situated all over the world. Their Experimental evaluation shows that the method increases accuracy of recommendation of Web services considerably. Anand Shanker Tewari, Kumari Priyanka (2014) [10] designed online book recommendation system for students reading textbooks. Motivation of this project was to develop recommendation system to the students according to their price range and publisher's name. Recommendation system is designed by combining the features of user based collaborative filtering and association rule mining. Recommendation is carried at offline mode and system stores recommendation in student's web profile. First system scans the buyer profile record find out the category of the book that the buyer has bought earlier such as Novels, educational book, Bibliography etc. Then it find out the subcategory of the book for example if at stage1 category of book is educational then subcategory would be Computer data structure, Theory of mechanics, operating systems etc. At stage third association rule is applied on transaction database and list of books is generated that buyer can buy it afterword's at stage 4 user based collaborative filtering technique is applied to the books found in stage 3 and top N recommendation are provided to targeted buyer. Jyoti Gupta and Jayant Gadge (2014) [11] used Collaborative Filtering algorithms to provide recommendations to a user. The recommendations are given based on the ratings of other users in the system. In their proposed work they have given solutions to overcome the Problems faced by collaborative filtering algorithms for instance sparsity, scalability, and cold start. In the proposed outline, item based collaborative filtering prediction is joined with prediction by means of demographics based cluster of user in an adaptive weighted scheme. Item similarity and user clusters are computed offline, which makes the solution scalable. A new user is added to the cluster with the nearest centroid based on demographics and can be given immediate predictions based on the user cluster. An adaptive weighting scheme will improve the quality of predictions by giving more weight to prediction using demographics based user clusters when the number of ratings available for a user are less, for example in cases when the user is new, and increasing weight to item based collaborative filtering as the number of ratings available for a user increases. This will result in a solution that is scalable while addressing user cold start. Anand Shanker Tewari, Abhay Kumar and Asim Gopal Barman (2014)[14] have designed recommendation system to recommend books to the buyer that suits their interests. The working of the system is divided into different stages. Stage1 finds out the category of the book that the buyer has bought earlier like novel, science, engineering etc. from the buyer's web profile. Stage2 finds out the subcategory of the book if there is any in the stage1 found category.stage3 performs content based filtering of category /subcategory found in stage1 and 2, to find out the books that are much similar to the books that the buyer has bought earlier based on the books overview content from the buyers past history record. At stage 4 based on the result of stage3 item based collaborative filtering is done to find out the list of books in the descending order of recommendations. In this step, the system actually evaluates quality of the recommending books based on the rating given to those books by the other buyers. From the book transaction database is used to find all those transactions whose category and sub category (if there is any) is same as found in stage1 and stage2. Stage5 applies association rule on those transactions and find out the books that the buyer can buy afterward. Stage6 discovers the intersection of the result of stage4 and 5. The intersection result is arranged in the descending order of recommendations as given by the stage4.Outcome of the stage 6 is the final recommendations for the buyer. All these steps are performed when the buyer is offline and the results are stored in the buyer's web profile. When the buyer comes online next time the recommendations will be generated automatically. ZhichaoQuan (2013) [12] in his research made use of user personality for the improvement of the user model and proposed two personality-based collaborative filtering recommendations are: first one is to calculate user similarity from the user personality point of view and choose nearest neighbor, and then produces recommendation; second one is using the personality-item rating matrix, and after that create recommendation for target users. Because of the stability of people's character, recommendation based on users' character is also relatively stable. This avoids the problem of frequently update of traditional user models, thus greatly reducing the cost of calculation. If user personality can be attained well, there will be no more cold start problems. The number of items of traditional collaborative filtering recommendation is too large, while personality factors in the proposed method is in small quantity and will not change as time goes by, which alleviate the sparsity problem to a certain extent. The weakness of this method is that: before recommendation, one must get users' characters, which would increase the work of the recommendation system. If there is no best way to obtain the character of the new users, "cold start" problem will occur. So when a new user is added, if the user character is not accurately obtained, their interests should be taken advantage of to discover their characters, and then make recommendation to users based on character traits.

5. CONCLUSION

Recommender System has made tremendous increase over the last decade. Collaborative filtering technique provides a solution to recommender system. In collaborative filtering the recommendations are given to the new user based on the preferences made by group of user who has similar tastes. In this paper, Collaborative filtering tasks, their main challenges and their possible solutions are discussed. This paper also covers major categories of CF techniques: memory-based, model-based, with examples for representative algorithms of each category, and analysis of their predictive performance and their ability to address the challenges. The study of various research papers and their results provides the fact that collaborative filtering is one of the most widely used filtering technique however it suffers from sparsity, accuracy, scalability, cold start etc.

REFERENCES

[1] Sarwar, B., Karypis, G., Konstan, J.A., Riedl, J., “Item-Based Collaborative Filtering Recommendation Algorithms”, Proceedings of the 10th international conference on World WideWeb. 2001, Hong Kong. ACM Press p. 285-295.

[2] http://en.wikipedia.org/wiki/Collaborative_filtering.

[3] Linden G. Smith B., York J., “Amazon.Com Recommendations: Item-To-Item Collaborative Filtering”, Internet Computing, IEEE, 2003.

[4] Francesco Ricci, Lior Rokach and Bracha Shapira, “Introduction to Recommender Systems Handbook”, Springer, 2011.

[5] L. H. Ungar and D. P. Foster “Clustering Methods For Collaborative Filtering”, in Proceedings of the Workshop on Recommendation Systems, AAAI Press, 1998.

[6] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “Grouplens: an open architecture for collaborative filtering of netnews”, in proceedings of the ACM Conference on Computer Supported Cooperative Work, pp. 175-186, New York, NY, USA, 1994.

[7] G.Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: a survey of the state-of-the art and possible extensions”,IEEE Transactions on Knowledgeand Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[8] http://en.wikipedia.org/wiki/Recommender_system

[9] Yechun Jiang, Jianxun Liu, Mingdong Tang and Xiaoqing (Frank)Liu ,“An Effective Web Service Recommendation Method based on Personalized Collaborative Filtering”, IEEE International Conference on Web Services, 2011.

[10] Anand Shanker Tewari and Kumari Priyanka ,“Book Recommendation System Based on Collaborative Filtering and Association Rule Mining for College Students”, IEEE International Conference on Contemporary Computing and Informatics, 2014.

[11] Jyoti Gupta and JayantGadge , “A Framework for a Recommendation System Based On Collaborative Filtering and Demographics”, IEEE International Conference on Circuits, Systems, Communication and Information Technology Applications, 2014.

[12] ZhichaoQuan , “Collaborative filtering recommendation based on user personality”, IEEE International Conference on Information Management, Innovation Management and Industrial Engineering, 2013.

[13] K.Yu, A. Schwaighofer, V. Tresp, X. Xu, and H.-P. Kriegel," Probabilistic memory-based collaborative filtering", IEEE Transactions on Knowledge and Data Engineering, vol. 16, no.1, pp. 56–69, 2004.

[14] Anand ShankerTewari , Abhay Kumar and Asim Gopal Barman, "Book Recommendation System based on Combine Features of Content Based Filtering, Collaborative Filtering and Association Rule Mining", IEEE International Advance Computing Conference, 2014.

[15] Su, Xiaoyuan, and Taghi M. Khoshgoftaar. "A survey of collaborative filtering techniques." Advances in Artificial Intelligence 2009 (2009): 4.

[16] K. Miyahara and M. J. Pazzani."Improvement of collaborative filtering with the simple Bayesian classifier",Information Processing Society of Japan, vol. 43, no. 11, 2002.

[17] B. Shen, X. Su, R. Greiner, P. Musilek, and C. Cheng, " Discriminative parameter learning of general Bayesian network classifiers", in Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, pp. 296–305,Sacramento, Calif, USA, November 2003.

[18] Breese. J. S, Heckerman. D, Kadie. C., "Empirical Analysis of Predictive Algorithms for Collaborative Filtering." Microsoft Research, 1998.

[19] J. Pearl, "Probabilistic Reasoning in Intelligent Systems", Networks of Plausible Inference, Morgan Kaufmann, San Francisco, Calif, USA, 1988.