RESOLVING BIAS IN RECOMMENDER SYSTEMS

[1] S.Ananthi, [2] M.Bharathi, [3] S.Anand

[1]Assistant Professor, [2]Assistant Professor, [3] Associate software engineer
[1]Computer Science and Engineering
[1]Sri Sai Ram Institute of Technology, Chennai, India

Abstract : **— Recommender systems or recommendation systems are a subclass of information filtering system that seek to predict the "rating" or "preference" that a user would give to an item. Based on your ratings, it generates personalized predictions. There is a huge load of incoming data which causes Bias which gives rise to Simpson's paradox. The recommendation system incorporates an algorithm that removes bias from the recommendation algorithms. One of the largest biases in this kind of data is due to the phenomenon of Simpson's paradox. "Simpson's Warning: less conditioning is most likely to lead to serious bias when Simpson's Paradox appears."This data without proper pooling can lead to incorrect recommendations and even ignored items that are potential recommendations. Hence, this could be solved by a technique called multi-resolution which is a method of pooling this incorrect data accurately.**

Keywords— **Hybrid Recommender System, Bias, Similarity, Filtering algorithms, Resolving bias in Recommender Systems.**

I. INTRODUCTION

The recommender systems are software applications, subclass of information filtering systems that analyzes available data to make suggestions that are interesting to users, such as books, movies or songs, among other possibilities. Recommendation systems seek to predict the "ratings" or "preferences" that a user would give to an item. Recommender systems selects the most suitable items from the abundant of choices available to them. The idea that led to the development of recommender systems was that, we people often depend on the suggestions of our peers for trying something new, say for before buying a smart phones, a laptops, before going for a movie, before going to a new restaurant and even before visiting a doctor. We have numerous recommendation systems developed for various fields, using different recommendation approaches.

Over several years, researchers have developed many recommendation engines for almost many of the domain like social-networking sites, entertainment, content based sites (e-learning, books or articles recommendation, e-filtering etc.), e-commerce, tourism, match-making etc, all dealing with real-world. We can classify recommender systems on the basis of their application areas: Entertainment, Social-networking based, Content-based, Services-based and E-commerce.

Basically recommendation systems use several approaches for the process of providing recommendations for the users. There are several approaches that recommendation systems use which includes Content Based Filtering (CBF), Collaborative Filtering (CF), and Hybrid systems which are the most frequently used approaches. However there are several more approaches like Demographic systems, Community Based systems, Knowledge Based systems. However we concentrate on Content Based, Collaborative, and Hybrid approaches for resolving limitation of Recommendation systems.

The aim of this paper is to acknowledge the effect of bias in the recommendation system, which is less considered as factor that degrades the performance of recommender systems. The efficiency of the recommendation systems can be improved to an extent if the problem of bias is considered and is resolved by the Hybridization techniques. The choice paradox, information overload, and bias resolved recommender systems can provide the user with increased efficiency of preferences and choices. The proposed system uses multi- resolution techniques, to classify the preferences and ratings. The explicit rating bias that occurs in recommender system as result of user rating is resolved by proposed system. Most recommendation models consist of building a user-by-item matrix with some sort of "interaction" number in each cell. If one includes the numerical ratings that users give items, then this is called an explicit feedback model. Alternatively, one may include implicit feedback which are actions by a user that signify a positive or negative preference for a given item (such as viewing the item online). These two scenarios often must be treated differently.

II. Collaborative Filtering

Collaborative filtering is the process of making automated predictions and suggestions about user's interest by collecting preferences from all or many of the users. A collaborative system filters information by recommendations of other users. The basic idea behind the collaborative filtering is that "Users who agreed in their evaluation of certain items in the past are likely to agree again in future". Collaborative filtering is referred to as people-to-people correlation. Basic concept of collaborative approach is that two or more individuals sharing some similar interests in one area tend to get inclined towards similar items or products from some other area too. The similarity between the users can be figured out on their browsing characteristics (click- through rate), browsing pattern and ratings (explicit, implicit).

The concept of collaborative filtering can be easily understood with the help of a simple example: consider Face Book. You can always see “people you may know” option on your home page with multiple people in the list. So the basic criteria behind making those suggestions are based on the concept of recommender systems only. The suggestions are filtered out on multiple parameters like: number of mutual friends you have with that individual, number of similar pages you both have liked or groups in common and also common places you have been to or you belong to. The approach used here is Collaborative filtering i.e. if a person x and you have a number of friends in common than the chances are that you two may also know each other. Hence it is called people to people co-relation.

In User-Based approach, for active user if ratings and user id is available, a group of similar users with same interest in terms of past ratings is found out and their ratings are used to predict what might interest the active user. In Item-Based approach it considers the similarity between the rating patterns of the items. For two items having similar users who like and dislike them, they are considered similar and the users thus are anticipated to have similar interests for similar items. It is similar to content based filtering in terms of overall structure, except the fact that here the similarity between the items is found based on the user ratings pattern and not on the item.

II. Content based filtering

Content based systems recommend items similar to items that user liked in past. Content based systems are for providing personalized suggestions for particular users. Content based systems are based on the concept . The basic idea behind these systems is to recommend items or products to a particular user, which are similar to the ones that user has already liked in the past. The similarity between two or more items can be calculated based on their similar features. To understand better consider the example discussed above, whenever you watch any video on Face book like the ones posted on food lover’s page, after you finish watching it, you get links of similar videos on your home page. Also, when you like some page, you get suggestions of pages similar to them.

Content based filtering i.e. if you like an ABC item from some particular category, it is likely that you might like any other item XYZ similar to it (from the same category or some category similar to it).There certain limitations in both of the above approaches that can be solved to an extent by combining the features of both the systems. So we go in for Hybrid Recommendation System.

IV. Hybrid approaches

A Hybrid system is one which combines the features of both Content based and Collaborative system. The Hybrid systems combine multiple recommendation techniques together to produce its output. Several studies and experiments show that performance evaluation of hybrid recommendations to be better than the collaborative and content based systems. A new system without collaborative data, the collaborative filtering approach cannot generate predictions. A system without content information, the content-based approach will fail to recommend. To overcome these principal limitations, hybrid systems have been developed.

V. Existing System

The hybrid recommender systems in use are subjected to certain limitations that result in the degradation of efficiency and performance. The challenges and problems in the recommender systems are explained below.

Cold start:

The problem occurs at early state of recommender systems when the information about an item or product available is less. The content based systems will behave poorly if product information available is less. On other hand if there are no people behavior history then the collaborative systems do not work efficiently.

Sparsity:

It is common in e-commerce and other domains that people usually purchase or rate relatively few items compared with the total number of items. That leads to a sparse users-items representation matrix and, therefore, inability to locate neighbors or derive common behavior patterns, and, the final result is low-quality recommendations.

Synonyms:

If single item is represented by two or more different words the recommendation systems consider that item as two or more different items and recommend to the user. This increases the number of recommendations which causes the same product to be recommended more than once.

Scalability:

It is another important concern in Recommender sytems. As ratings database grows, the performance decreases. It is beneficial to try to make systems, which can handle large amounts of data and produce accurate recommendations quickly.

VI. Proposed System

In the proposed system the modifications are done to existing hybrid recommendation system by implementing changes in algorithm that resolves the problem of bias and choice overload in the existing hybrid systems. For the same data set the recommendations suggested by the proposed system tends to show better recommendations. The existing system is remodeled into a hybrid recommender system with bias factor being resolved. Proposed system incorporates multi-resolution technique that resolves the bias to considerable extent when deployed to the hybrid recommender systems.

i) Hybrid recommender system
ii) Bias resolved hybrid system
iii) Performance evaluation metrics
iv) System Architecture

i) Hybrid Recommender System

The hybrid recommender system is the combined form of system with functionality of both content based and collaborative systems. The system used in the project incorporates the hybridization techniques for the process of preferring and suggesting the user with recommendations. The system accepts the input from user as search query and results for the query are the set of recommendations that the user needs. The system provides preferences to user but they are subjected to bias which will not be resolved by this system. This system incorporates both the content based and collaborative techniques but has some efficiency issues.

This step computes the rating predictions of all unrated apps by an active user. The predicted rating of active user *a* on a target app *x* , $P^{IT}_{a,x}$ is calculated using the weighted sum of deviations from the mean app ratings approach

$$P^{IT}_{a,x} = \bar{r}_x + \frac{\sum_{y=1}^{N^{IT}} Itrust_{x,y}\,(r_{a,y} - \bar{r}_y)}{\sum_{y=1}^{N^{IT}} Itrust_{x,y})} \quad (1)$$

$$(\text{if } Itrust_{x,y} \neq 0)$$

$$P^{IT}_{a,x} = \bar{r}_x + \frac{\sum_{y=1}^{N^{IT}} IR_{a,y}\,(r_{a,y} - \bar{r}_y)}{\sum_{y=1}^{N^{IT}} IR_{a,y})} \quad (2)$$

This module to combine the prediction values of the user-based trust and the app-based trust approaches. The hybrid prediction value $HP_{a,x}$ take into account all possible ways to obtain a rating prediction value for an active user *a* who has not rated the target app *x*. The weighted harmonic mean method is used to ensure that a high total prediction rating value is obtained only if prediction rating values of both the implicit user-based and the implicit app-based trust approaches are high.

$$HP_{a,x} = 0, \qquad (\text{if } P^{UT}_{a,x} = 0 \text{ and } P^{IT}_{a,x} = 0)$$
$$HP_{a,x} = P^{UT}_{a,x}, \qquad (\text{if } P^{UT}_{a,x} \neq 0 \text{ and } P^{IT}_{a,x} = 0)$$
$$HP_{a,x} = P^{IT}_{a,x}, \qquad (\text{if } P^{UT}_{a,x} = 0 \text{ and } P^{IT}_{a,x} \neq 0)$$
$$HP_{a,x} = \frac{2 * P^{UT}_{a,x} * P^{IT}_{a,x}}{P^{UT}_{a,x} + P^{UT}_{a,x}}, \qquad (\text{if } P^{UT}_{a,x} \neq 0 \text{ and } P^{IT}_{a,x} \neq 0)$$

$P^{UT}_{a,x}$ User-based trust predicted rating value of the active user a *on* a target app *x*

$P^{IT}_{a,x}$ App-based trust predicted rating value of the active user *a on* a target app *x*

ii) Bias Resolved Hybrid System

The bias resolved hybrid recommender system is the modification of the above recommender system. The both systems implement the content based and collaborative filtering with hybridization techniques. But the previous system does not account for resolving the bias that occur in the suggestions and rating. This problem of bias and choices are overcome by this bias resolved hybrid systems. These systems have been incorporated with multi-resolution techniques in the hybrid recommendation algorithms that overcome the problems of bias in system.

iii) Performance Evaluation Metrics

The performance metrics are used to evaluate the performance of the recommendation systems. Here we measure the performance of the both system, the hybrid recommender system and the bias resolved hybrid system. The performance of the systems can measured and represented as visualize function between two systems. Several of the performance evaluation metrics for visualization of the performance we can use Mean Average Precisions or Normalized Discounted Cumulative Gain.

iv) System Architecture

The architecture of the proposed model can be represented as a hybrid recommendation system that is a combined model of content based and collaborative approach. This tells how the system processes the information source and user profiles, so that they can recommend items based on the process. The three main components of the system include the Content analyzer, Profile learner, Filtering component.

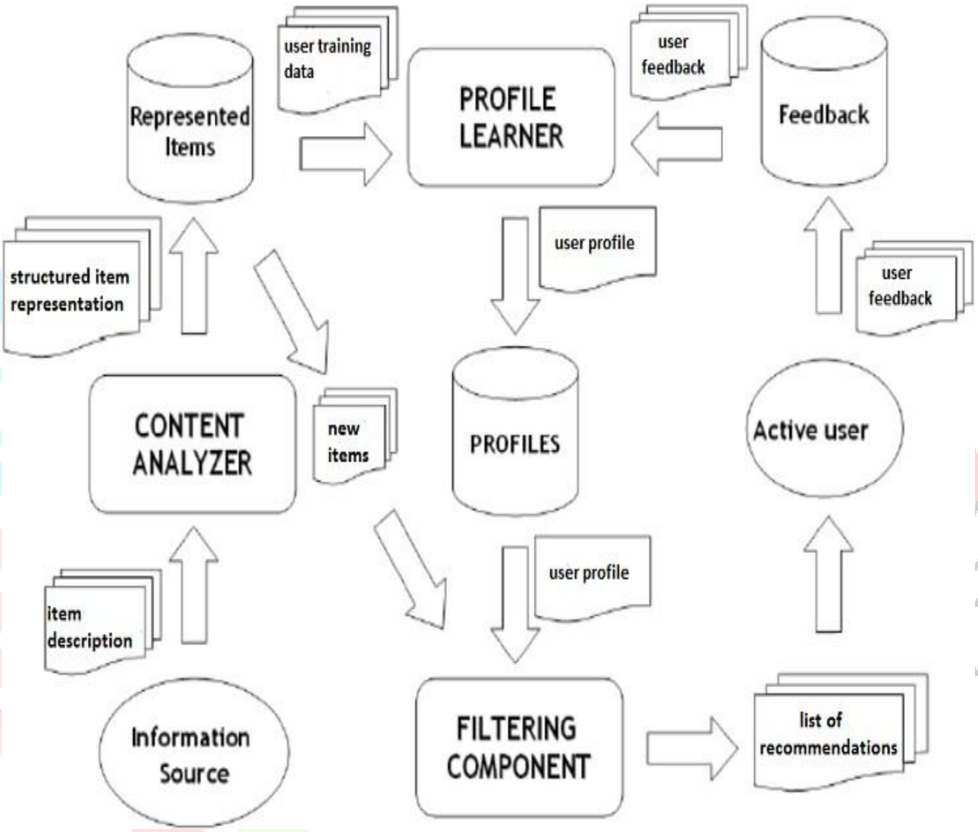

FIG 6.1 SYSTEM ARCHITECTURE

VII. Conclusion and Future enhancement

The recommendation systems can enhanced with bias resolving algorithms that reduces bias in the recommendations that are being recommended to the user. The basic hybrid recommendation systems are prone bias and have to be resolved. The resolved bias reduces the choice over load, and provides the user with best and the reduced number of recommendations or suggestions. This reduces the user's time for searching the items from the list of all items available. This can be implemented as an enhancement in all of the recommendation engines that existing real-time to help user in choosing form a comparatively smaller subset of items available. The bias resolved hybrid recommendation systems performance measure can be evaluated to the performance of the normal hybrid system and the result of the performance evaluation denotes that bias resolved systems have better efficiency than the normal systems. The advantage of our project is that the user can feel easy in choosing the items from recommended list and can have a bias free recommended list for searching the items. The hybrid systems become more faster and efficient if the bias and information overload is resolved.

In future more detailed information of the items can be recommended by the recommender systems, and more efficient way of handling the problem like choice overload and bias can be made. In future the recommendation systems incorporate machine learning algorithms and collaborate user data with server data to provide suggestions at a faster rate. The several more hybridization techniques can be implemented to the hybrid recommendation systems for the user to easily select from available items. Explicit rating bias can be resolved by hybridization techniques. Steps will be taken to make the algorithm run with lesser resources and faster speed to improve performance.

VIII. References

[1] J Bobadilla, F Ortega, A Hernando, 2013 proposed"Recommender Systems Survey"

[2] Gediminas Adomavicius, Jesse Bockstedt, Shawn Curley, and Jingjing Zhang , 2014 proposed "De- Biasing User Preference Ratings in Recommender Systems"

[3] Sanjay Krishnan, Jay Patel, Michael J. Franklin, Ken Goldberg, 2014 proposed "Social Influence Bias in Recommender Systems"

[4] Asela Gunawardana, Christopher Meek , 2009 proposed "A Unified Approach to Building Hybrid Recommender Systems "

[5] Adomavicius, G., Bockstedt, J., Curley, S., and Zhang, J. 2013. "Do Recommender Systems Manipulate Consumer Preferences? A Study of Anchoring Effects," Information Systems Research,24(4)

[6] Bell, R.M., and Koren, Y. 2007. "Improved NeighborhoodBased Collaborative Filtering," KDDCup'07, San Jose, CA, USA, 7-14.

[7] Cosley, D., Lam, S., Albert, I., Konstan, J.A., andRiedl, J. 2003. "Is Seeing Believing? How Recommender Interfaces Affect Users' Opinions," CHI 2003 Conference, Fort Lauderdale FL.

[8] Koren, Y., Bell, R., and Volinsky, C. 2009. "Matrix Factorization Techniques for Recommender Systems," IEEE CS, 42, 30-37.

[9] Sarwar, B., Karypis, G., Konstan, J.A., and Riedl, J. 2001. "Item-Based Collaborative Filtering Recommendation Algorithms," Int'l WWW Conference, Hong Kong, 285 - 295.

[10] Bardul M. Sarwar, George Karypis, Joseph A. Konstan, and John T. Riedl, "Analysis of recommendation algorithms for ecommerce," in Electronic Commerce, 2000.

[11] Mark Claypool, Anuja Gokhale, Tim Miranda, Pavel Murnikov, Dmitry Netes, and Matthew Sartin, "Combining content-based and collaborative filters in an online newspaper," in ACM SIGIR Workshop on Recommender Systems-Implementation and Evaluation, Berkeley, CA, 1999.

[12] Daniel Billsus and Michael J. Pazzani, "Learning collaborative filters," in 15th International Conference on Machine Learning, Madison, WI, 1998.

[13] Shambour, Qusai, and Jie Lu. "An effective recommender system by unifying user and item trust information for B2B applications." Journal of Computer and System Sciences 81.7 (2015): 1110-1126.