“A SURVEY FOR RECOMMENDER SYSTEM” -WITH LATEST CHALLENGES

Jitali Patel
Assistant Professor at Department of Computer Science and Engineering, Nirma University

ABSTRACT

Today people are flooded with many options on internet. Recommender system collect information about the item according to the preferences of the users. Recommender system are successfully implemented in different e-commerce setting. The objective of this paper is to show various techniques being used for recommendation and to discuss various challenges especially for the web media sites.

Keywords : collaborative filtering, content based filtering, hybrid filtering, personalized recommendation, recommender system

I. INTRODUCTION

The idea of recommender system generally goes out of the idea of information reuse and persistent preferences. And the idea of recommender system does not began with computers and technology it is an idea that you confined in the ants and caveman and also other creatures. So the idea of recommender system comes from social navigation, following in the footstep of others to find what you want.

Recommender systems support users in personalized way for the identification of product based on the history of the user that can be useful or interesting in the large space of possible product [1]. Recommender system is an information filtering technology, commonly used on e-commerce Web sites to present information on items and products that are likely to be of interest to the reader. In presenting the recommendations, the recommender system uses details of the registered users’ profile, opinions and habits of their whole community of users and compares the information to reference characteristics to present the recommendations [1]. As a research discipline, recommender systems has been established in the early 1990's and since then it has shown enormous growth in terms of algorithmic developments as well as in terms of deployed applications. Practical experiences from the successful deployment of recommendation technologies in e-commerce contexts (e.g., amazon.com and netix.com) contributed to the development of recommenders in new application domains.

Recommender Systems are the instances of Personalization software. Personalization concerns with the adapting to the individual needs, interest and preferences of each user. It includes:

- Recommending (suggesting list)
- Filtering (Email filters)
- Predicting (Predicted rating)

At vary basic level there are two types of recommendation techniques, personalized recommendation and non-personalized recommendation system. In this paper before discussing the major challenges of recommendation system, the ample survey of recommendation approaches is given. The discussion of various approaches and their limitations in a proper manner thereby provides the future research possibilities in recommendation system.

There exist three basic recommendation approaches for personalized recommendation at the algorithmic level. First is a content-based filtering, Content based Filtering characterizes the affinity of users to certain features (content, metadata) of their preferred items. The filtering is done on the basis of the attributes of the item. For example, in the case of movie, who are the actor, director, genre and so forth? Machine Learning Algorithm is used to induce a profile of the user's preferences from examples based on a feature description of content. There are various other classification technologies under this area to decide whether the item is applicable to the user.

Second is a Collaborative Filtering exploits similar consumption and preferences pattern between users. It maintains the database of many users' ratings of variety of items. For a given user, it finds other similar users whose ratings strongly correlate with the current user. Recommending items rated highly by these similar users, but not rated by the current user. Almost all existing commercial recommenders use this approach (e.g. Amazon) [1]. And again there are two variation with this technique, user-user collaborative technique and item-item collaborative technique.

And third is a Hybrid Filtering is combination of both Content based Filtering and Collaborative Filtering. Therefore Hybrid Filtering enjoys the advantages of both Content based Filtering as well as Collaborative Filtering.

II. METHODOLOGY

Any recommendation system requires a concrete model and for that model we have their basic concept that every recommender needs. As shown in fig 1. There is a notion of users, users are the people in our system the people who have preferences for items and who may the source of data in the case of collaborative recommendation system. Items, these are the things we choosing to recommend and there are ratings which are the things that express an opinion in our system. And more broadly when we take the users and items together we tend think of having a community that expresses a space in which all these opinion make sense. look at the following figure each user may have set of attributes for example demographic would be a user attribute and for each user we may have a set of user model such as their preference of type of book they like to read and for each item we may have set of item attribute or properties for example who is the author of that book and where user meets item is in the space of ratings. Normally we think of a preference expressed as a user likes a particular item or user purchase that item or user may express an opinion over a things that is not an individual item.

III. NON-PERSONALIZED RECOMMENDATION SYSTEM

First type of recommender system is non personalized recommender system is involve summary statistics or in some cases product association. This use external data from community like this is a best seller or a most popular video or it may provide summary of community ratings, how much somebody like a restaurant or a summary of community ratings that turns into a list like what is the best hotel in town. The algorithms used to implement non personalized recommender are aggregated opinion recommenders and Basic product association recommender. The example of aggregated opinion recommender is Netflix.com. Look at the framework of recommender system for the understanding of aggregated opinion recommender. The simplest example we may have is would be the summary stat in the rating for a movie site like Netflix. Each item would be a movie, each user is visitor to a site when a visitor to a site enter rating for a movie that ratings might be four star or a five star, because this is just a summary of ratings for this purpose we don't need any attributes we don't need any model, when somebody comes and says how good this movie we go to the ratings for this movie pull out all the ratings that match with this movie and report the average may be three star or may be four and half star and this the simplest way to do this. Rating table is a matrix where we have one dimension as a user and one dimension as an item.

Table 1. Non personalized recommendation

	Item1	Item2	Item3
User1	2	3	3
User2	1	4.5	3
User3	4.5	2	3

The problem with this approach is average may be misleading. Context in recommender can be quite important. You want to know what a recommendation is being used for what situation. Consider that you are at a restaurant and you are ordering an ice-cream Sunday and you want to know which sauce I want to put on my Sunday ice cream now if you go to the waiter and says what is the most popular sauce and waiter says catch up is the most popular sauce and you might not find to be the most useful recommendation for your Sunday this leads to the concept of product association recommenders. Lots of sites recommend the item that frequently brought together. So for this second type of non-personalized recommendation concept of association rule mining is used. This non personalized recommendation still face challenges in a clustered diverse population. Clustered here is that interest are not just uniform ally distributed among everyone it may be true that 95 percent of American eats bananas but that probably doesn't show evenly among different cultural groups , from some cultural group it may very close to 100 percent in other perhaps fewer than half of the people eat bananas and in overall average is not going to solve that problem we can solve this by different community, we can say this is the average of chines Americans and what they brought this is the average of iris Americans and what they brought or that point we may say it's time to start thinking about personalization and start saying what's the average of people like our current consumer.

IV. PERSONALIZED RECOMMENDATION TECHNIQUES

1. Content-based filtering

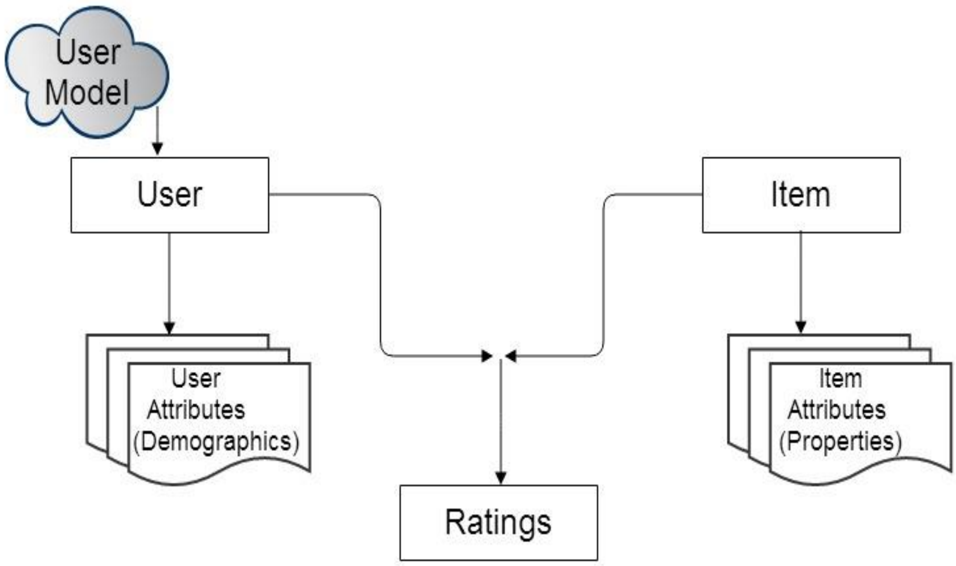

Fig 1. Framework of recommendation system

Look at the framework of recommendation system in fig 1

In content based filtering user rate items and from that we build a model of user preferences against the item attribute. For example movies will have some important properties like genre, actor/actress director and when user submit a rating for a particular movie, I say krish is a five star, that updates a model where the model is not about the which movies I like but which properties of movie I like and inside that model there is going to be a table, frequently referred to as keyword vector or a test vector. And every time I rate a movie it can update this model and when the times comes for recommendation, when I look for a particular movie that comes in, it will map how much this movie have each of this attribute, how much is that action, how much science fiction and then stronger it has the attribute the more I look at whether that something user like or dislike.

The application where the content based filtering would be applicable is the *personalized news feeds*, because news is coming out so quickly that you may not wait for the other people's opinion. Personalize news for each person case, whether the person is interested in political news, less interested in Bollywood news[3].

2. Collaborative Filtering system

Use opinions of others to recommend rather that attribute of data to recommend. In collaborative filtering lots of items and lots of user and it's filled in very sparsely. If the matrix is close to full we don't need recommender system we know everybody's opinion

Table 2. content-based filtering

	Item1	Item2	Item3	Item4	Item5
User1			2		4
User2		2	3		
User3	1			5	
User4	5	2			
User5			4		3
User6				1	

There are two variation with collaborative filtering, User-User collaborative filtering and Item-Item collaborative filtering.

In User-User collaborative filtering select neighborhood of similar-taste and use their opinion. In Item-Item collaborative filtering you establish relationship among items via ratings and use those items and rating for recommendation. Issue with User-User collaborative filtering is *sparsity*, with large item set, small number of ratings too often there are points where no recommendation can be made (for a user, for an item, to a set of users etc..). Solution to this issue can be found using filterbot. Filterbot is nothing but a combination of content based technique with collaborative by creating agent that rate everything based on person's profile. Another issue with user-user collaborative filtering is computational performance, with millions of users computing all pair's correlation is expensive. Even incremental approach is expensive. User's profiles would change very quickly and require updating in real time to keep user's happy. So **item-item** similarity is fairly stable. The idea is 'item similarity is a route to computing a prediction of a user's item preferences. We can compute similarity between pairs of items using correlation between rating vectors. For example look at the following table

Table 3. collaborative filtering system

	Item1	Item2	Item3	Item4
ann		5	3	1
Bob	4	5	3	
chang	2	1		5
elice		2	3	
dhgg	5	1	2	4

We can see from the table that item2 and item4 is having negative correlation while item2 and item4 are strongly correlated because higher value corresponds to higher value and lower one corresponds to lower one. So we want to predict the missing value for chang for item number 3 and there is a very strong relation between item2 and item3 and chang didn't like item2 so we can predict that chang didn't like item3. If we have more than one column with correlation than it takes the average of all[4].

V. LATEST RESEARCH CHALLENGES

1. Inferring implicit interactions and satisfaction:

Generally, the matrix consists of <user-item-rating> triplet which is based on the external ratings.

In many recommendation settings, we only know which items are consumed by the users and not whether they liked or disliked i.e. no explicit rating data is maintained. This is called Binary consumption model. But in this model, the question that comes what about the item that the user did not consume. Was the user aware of the items he did not consume? This might not be the problem in the case of movies, books and the like items as the user was exposed to them anyway. But in the case of media sites the items are ephemeral and hence if the user is not exposed to them by the recommender system the user will end up selecting the best out of the little that was shown. Presentation bias obscures the true taste.

The presentation bias can be accounted if we treat them as:

Seen and not selected ≠ not seen and not selected

Also, the "skipped" results can be negatively interpreted. It can be inferred that the user might not have liked the item.

2. Layout of recommendation modules:

Vertical Layouts: This type of interpretation is easy. But in other cases it might be difficult where there is different stress on different items.

2D Layouts: how would the different users skip in 2D layouts and will their behaviour be same or different is a question.

Horizontal Layouts: Whether the people scan from right or left or middle?

Tabbed Layouts: The items will be seen by them in the same order or not?

Exploring these research area will help in knowing how the users express negative feedbacks on the item. Also, in vertical layout the issues that may come up due to multiple presentation formats and would the user skip or select the salient item.

3. Inferring the satisfaction:

For example: In web pages what happens after the initial click (Did they like the content or not), in online video what happens after the pressing of play key (Did they rewind and repeat some areas or they just went through the item for some seconds and left), in TV what happens after zapping platforms.

Sometimes the scenario is that we are not sure about whether the user has consumed the item. For example, in TV domain as the time passes after zapping of the channel there are chances that the user might have dozed off and hence nobody is watching the program despite of the TV being on. In this case the binary consumption model also fails.

4. **Personalization vs. Perspective**

Web media sites often display the links to additional stories on each article page. It may come from The matching of article's content, The popular items, Based on the item previously consumed by the user Based on the item consumed by the user's friends. The challenge here is when the user is reading a specific page, how the unified list is created mixing personalization and perspective. Ignoring context stories might create offending recommendations. For example, If the user is currently reading some note related to the current natural disaster and he has penchant for say Dan Brown novels then it may not be appropriate to recommend him articles related to disaster every time.

5. **Repeated consumption and repeated recommendation : Diversity**

In the case of books and news stories, the user would consume the item only once. On the contrast, in the case of movies and songs the user would like to watch the same movie/song multiple times.so the challenge is When and how frequently it is reasonable to recommend the item that is already consumed. For example, if user saw a movie yesterday he might not want to watch the same movie today but if the user had listened song yesterday he might like to listen the same song today. When should we stop recommending, if the user has stopped to act upon the item? This means that in addition to tracking the aggregated consumption to-date, it may need to track consumption timelines and recommendation history

Diversity:

- How diverse is the recommendation to the given user at time t?
- How diverse are the recommendations across all users at time t?

  Indication of aggressiveness of personalization and deviation from popularity baselines.

- How diverse are the recommendations to the user u over time? The system should not recommend the same item over and over again.

6. **Incremental collaborative filtering:**

Live system often cannot afford to recomputed recommendation regularly over the entire history. Issue here is collaborative filtering models do not easily lend themselves to faithful incremental processing.

VI. CONCLUSION

Through Recommender System, we can achieve personalize user interest into E-Commerce and web media web sites. In this paper, we gain recommender system via different techniques like Content based Filtering, Collaborative Filtering and Hybrid filtering. This paper describes the various latest challenges or issues specially for web media. This can be helpful for the people who are interested in this field.

VII. REFERENCES

[1] Alexander Felfernig, Michael Jeran, Gerald Ninaus, Florian Reinfrank, and Stefan Reiterer, "Toward the Next Generation of Recommender Systems: Applications and Research Challenges," Springer Link, vol. 24, pp. 81-98, 2013

[2] Gideon Dror, Noam Koenigstein, and Yehuda Koren, "Web-scale media recommendation system,"Procedding of IEEE, vol. 100,Issue:9 no. 5, pp. 2722-2736, May 2012.

[3] Lalita Sharma , Anju Gera, "A Survey of Recommendation System: Research Challenges," International Journal of Engineering Trends and Technology, vol. 4 Issue 5 2013.

[4] Zhongyun Ying, Zhurong Zhou Fengjiao Han and Guofeng Zhu "Research on personalized web page recommendation algorithm based on user context and collaborative filtering," IEEE Trans. Software Engineering and Service Science., pp. 220-224, May 2013.