Trends, problems and solutions of recommender system

4 authors:

Sarika Jain
National Institute of Technology, Kurukshetra
81 PUBLICATIONS **357** CITATIONS

Anjali Grover
Beuth Hochschule für Technik Berlin
1 PUBLICATION **24** CITATIONS

Praveen Singh Thakur
National Institute of Technology, Kurukshetra
1 PUBLICATION **24** CITATIONS

Sk Choudhary
National Institute of Technology, Kurukshetra
1 PUBLICATION **24** CITATIONS

Some of the authors of this publication are also working on these related projects:

Leveraging AI in Global Epidemics View project

A Multilingual Semantic Web Portal for Unconventional Emergencies View project

Trends, Problems And Solutions of Recommender System

Dr. Sarika Jain, Anjali Grover, Praveen Singh Thakur, Sourabh Kumar Choudhary

Department of Computer Application, National Institute of Technology Kurukshetra,
Kurukshetra, Haryana, 136119, India
jasarika@gmail.com, anjaligrover93@gmail.com, praveenthakur1301@gmail.com, ssouravchoudhary@gmail.com

Abstract—**In this era of web, we have a huge amount of information overloaded over Internet. It becomes a herculean task for the user to get the relevant information. To some extent, the problem is being solved by the search engines, but they do not provide the personalization of data. So, to further filter the information, we need a recommendation engine. In this paper, we have described the various web recommender systems in use by some popular web sites on the internet like Amazon.com, LinkedIn.com, and YouTube.com etc. Further, we have described the various approaches used in the various recommender systems such as Content based, Collaborative and Hybrid recommender system. At the end of this paper, we focus on some of the main challenges faced by the web recommender systems and analyze some techniques to overcome them.**

Keywords—Recommender System, Content Based, Recommender System, Collaborative Recommender System, Hybrid Recommender System

Abbreviations-- R-S (Recommender Systems),CF (Collaborative Filtering)

I. INTRODUCTION

It is mostly necessary to make choices without prior personal experience or knowledge about something. In our everyday life, we depend on recommendations given by other people either by word of mouth or reviews of general surveys. Suppose, many of the times you walk-up to the restaurant, may be you are there for the first or second time, since you don't have any idea what to order so you go through the menu and take some time to decide your order and in the meanwhile waiter comes to you and ask "what would you like to take?", and you ask for some more time but even after sometime you are still there and you haven't decided what to order. So in this case, you ask waiter to recommend something to you, then on the basis of provided recommendations, you give your preferences. Similarly, people often use recommender systems over web to make decisions for the items related to their choice. Recommendation systems are software tools and techniques whose goal is to make useful and sensible recommendations to a collection of users for items or products that might interest them. In other words, recommender system or recommendation systems belongs to a class of information filtering system that aims at predicting the 'preference' or 'rating' given to an itemby any user [1][2]. Recommender systems are implemented in a variety of applications and have become really useful in recent years. The most famous areas where the concept of recommender system is implemented are movies, music, news, books, social tags, products, restaurants, financial services, life insurance, persons (online dating), Facebook friends and Twitter followers. Recommendation algorithms are widely used on E-commerce websites [3] where they use information about a customer's interests as input and generate a list of recommended items.

However, in spite of all these advances the current generation of recommender systems still requires further necessary improvements to make recommendation approaches more effective and more applicable on a wide range of real-life applications.
In this paper, we first present a comprehensive survey on trends of recommender system, then we identify various cons of the different recommendation methods and discuss some initial approaches to extend their capabilities.

II. RECOMMENDER SYSTEM AT GROUND ZERO

Recommender system emerged as an independent research area in the mid 1990's when researchers started focusing on recommendation problem that explicitly depends on the rating method.

"Getting information off the internet is like taking a drink from a fire hydrant"[4].

As we have a huge amount of information overloaded over Internet. It becomes a herculean task for the user to get the relevant information. A recommender system provides a solution for this by allocating suggestions for the items to be of use to a user.Recommender system serves the "right" item to the users in an automated fashion to optimize long term business objectives. Automation comes through algorithms. Some of which we will discuss in the upcoming section of the paper. Moreover, using recommendation system enterprises can get to know customer's 'buying behaviors better' and develop efficient marketing strategies to attract different customers.
Let us take an example of the video recommendation system in use at YouTube, the world's most liked online video community. YouTube video recommendation system delivers personalized sets of videos to signed-in users based on their

past activity on the YouTube site(while recommendations are also available in a limited form to sign out users).

The set of recommended videos is made by using a user's personal activities (watched, favorite, liked videos).The primary goal for YouTube recommendations is to maintain user privacy and provide explicit control over personalized user data that back-end systems expose to boot. [5]

You can also find recommender systems in many of the websites you use every day. Includingthese well-known examples:

LinkedIn, the business-oriented social networking site, forms recommendations for peopleyou might know, jobs you might like, groups you might want to follow, or companies youmight be interested in. LinkedIn uses Apache Hadoop to build its extensive collaborative-filtering capabilities.

Amazon, the popular e-commerce site, uses its own recommender system. When youselect an item to purchase, Amazon recommends other items other users purchased basedon that original item. Amazon patented this behaviour, called *item-to-item collaborative filtering*[6].

III. BREEDSOFRECOMMENDERSYSTEM

Recommender Systems are primarily categorised on the basis of personalized recommendations and Non-personalized recommendations. Personalized recommendations are offered as ranked list of items. Personalized recommender systems are used by E-commerce sites to recommend products to their customers (based on past activities of the individual). Non-Personalized recommendations which are must simpler to generate and are normally featured in magazines or newspapers. Non–personalized recommender system recommends products to customer based on what other customers have said about the products on average. The recommendations are not dependent on the customers, so each customer gets the same recommendations.

Recommender Systems can be classified broadly into several categories depending on the information they use to recommend items:

✓ Content-based Filtering Systems: Uses information of active users and data about the items.

✓ Collaborative Filtering Systems: Uses information about a set of users and their relations with the item to provide recommendations to the active user.

✓ Demographic Filtering Systems: Uses demographic information such as age, gender, education, etc. of people for identifying types of user.

✓ Hybrid recommender Systems: By putting trustworthiness as a forward feature, it uses combination of Content-based and Collaborative filtering.

A. Content-based Filtering R-S

Content-based filtering recommender systems recommends items based on contents of items rather than other users rating of the system. Instead of using a user-to-item correlation and defining methodologies, they use item-to-item correlation for generating recommendations. Following steps are carried out in the process of generating recommendations:

- Gathering content data about the item (For example- title, author, cost etc. for the books are some of the common content information.

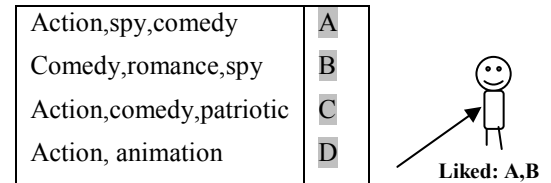- Process data and extract useful features and elements about its content.

Figure 1 – Content based approach

Advantages: Content based approach doesn't require data of other users and has capabilities of recommending items to user with unique taste. It avoids first rater problem.

Disadvantages: In content based filtering items are limited to their initial descriptions or features.

B. Collaborative Filtering R-S

In Collaborative filtering, recommendations are based on a few customers who are most similar to the active users. It fathoms the similarity of two users in various ways; one common method is to calculate the cosine of the angle between the two vectors: [7]

$$\cos(\vec{A}, \vec{B}) = \frac{\vec{A} \bullet \vec{B}}{\|\vec{A}\| * \|\vec{B}\|}$$

The collaborative filtering can be adapted with neighbourhood methods, whose focus is on relationship between the items or, alternatively between the users. They are:

I. User-based CF: For each user, compute correlation with other users. For each item, aggregate the rating of the users highly correlated with each user. Problem: Sparsity, easy to attack

II. Item-based CF: For each item, compute correlation with other items. For each user, aggregate his rating of the items highly correlated with each item

Advantages: Collaboration filtering approach doesn't need a representation of items in terms of features but it is based only

on the judgement of participating user community.

Disadvantages: The item can't be recommended to any user until and unless the item is either rated by another user(s) or correlated with other similar items.

C. Demographic Filtering Systems:

It uses pre-existing knowledge of demographic information about the users and their opinions for the recommended items as a basis for recommendations. Demographic systems are stereotypical, because they depend on the assumption that all users belonging to a certain demographic group have alike taste or preference.

Let us take an example, suppose Sourabh (male, age: 20) purchased items A and B from a website, Anjali (female, age: 25) purchased items A, B and D from the same website and Praveen (male, age:21) purchased items A and C from the same website. Now Sourabh wants to purchase an item from recommended items. He will get the recommendation for item c based on his demographic information because both Praveen and Sourabh lies in same age group and have purchased items of similar taste.

Advantages: It does not require history of user ratings that are required by collaborative and content-based techniques. This is a quick, easy and straight forward approach for making results based on few observations.

Disadvantages: Concerning the security and privacy issue, gathering of complete user information is impractical.

D. Hybrid Recommender Systems:

It is another category of recommender system that tries to overcome the limitations of the other approaches discussed before. It is a combination of two or more different recommendation techniques. The most popular hybrid approaches are those of content based and collaborative filtering. It use both item content and the ratings of all users [8].

E. Context-Aware Recommender System:

It is one of the most trending recommender systems these days. It helps in giving diverse and accurate recommendations to the user. The contextual information may include location of the user, Identity of people around, date, season, temperature etc.[10] The contextual information may be retrieved in a number of ways, including:

a. Explicitly i.e. gathering information by asking the direct questions from the user. For example, a website may recommend songs to a user by asking the current mood of the user.

b. Implicitly i.e. from the data or the environment,

c. Inferring (To conclude from evidence or by reasoning).

IV. CHALLENGES AND SOLUTIONS

The web recommender System suffers from many challenges such as Lack of Data, Changing Data, Changing User Preferences, Unpredictable Items, Scalability, Privacy protection. Some of them are:-

Cold-start,Privacy, Over-specialization, Scalability,Sparsity, Freshness etc.

Cold-start

Cold start problem can be classified into two categories, cold start of new items and cold start of new users. Cold start problem for an item occur when we don't have enough previous rating related to that item. Also, it is a bit difficult to recommend items to new users as the system don't have any information related to his past purchases or it might be possible that he has not rated any item yet so his taste is unknown to the system.

Scalability

As the numbers of users and items grows the system needs more resources in order to give the most accurate recommendations to the users. Most of resources are used in the purpose of determining users of similar tastes, and items with similar attributes. It is one of the problems found in collaborative filtering approach.

Sparsity

Suppose, you have an online shop that has a huge amount of users and items. If a user purchased few items from the shop and has rated any of them. Then, it will lead to the problem of Sparsity. Also, we can say that Sparsity is the problem of lack of knowledge.

Privacy

Privacy is also a big issue in context of demographic recommender systems. In order to give the most accurate recommendation to the user, the system must acquire the most appropriate information of user, including demographic data (age, sex, email-id, hobbies etc.), and data about the location of a particular user which may breach the privacy of the user.

Over-Specialization

This is one of the most common problems faced by the content-based recommendation system. A good recommender system must suggest diverse items which content-based system lacks. It gives nothing "surprised". It hinders the users from discovering something new and different. Users are recommended items they are already familiar with.

Freshness(Predictability):

One more problemgenerally faced by recommender systems these days is that of **predictability**. Even if the items

recommended to the user are diverse, it might be familiar to the user. For example, a system recommends best sellers only. The recommendations in this case are indeed diverse but the user may already know or be familiar with the recommended stuff.

V. SOLUTION

Cold-Start:

To curb the problem of cold-start, we can use the demographic information of the user from social networking sites or through the sign up page of the website. Also, we can use hybrid approach, i.e. to use collaborative filtering with demographic recommending approach to suggest items to a new user.

Sparsity:

The problem of Sparsity can be resolved using hybrid recommendation technique. Instead of using content based alone we can combine the content based and collaborative technique together which will result as a solution of Sparsity. The amount of information people have in common can be increased by using the attributes of an item instead of the item itself.

Overspecialization:

The problem of overspecialization can be overcome with the neighborhood based collaborative filtering technique. For the probabilistic neighborhood selection phase, we use a method for weighted sampling of k neighbors that takes into consideration the similarity levels between the user/items and candidate neighbors. [9]

VI. OUR PROPOSED APPROACH

1. As the new user login to the system ,system will collect all his demographic details and location based on his IP address through which we get to know about his zip-code also system will start recording his /her activities such as click made on the links ,time spent on the pages.
2. Since user is new to site so that it may fall for cold start problem, to solve that initially we recommend him the item that belongs to the dataset of most selling items in that particular area (zip-code). As the user click on link of any item a time session will be started to record how much time he has spent on that particular page. When the time spent by the user crosses a certain threshold time, rating of that particular item will be increased by some measures. Which solve the problem of Sparsity and result in efficient recommendations.

3. We would recommend our items in two sections:

I. In the first section, items would be recommended based on the most recent links visited by the user. At every click of the link by the user, we would store the item id, user id, item category, price range, and no. of clicks in our database. This would help us to recommend items according to the changing preferences of the user.

II. In the second section, we would recommend items based on the purchases done by the user in the past .We would store the purchasing history of the user in our database. So, that relevant and of user's interest could be recommended to it the next time as well.

4. This approach would be used when the user would use the site for the first time .Initially; we rate or provide weight to each and every item using a function.This would help us to overcome the cold start problem. And diverse items would be recommended to the user.

VII. CONCLUSION

Over the last decades recommender systems emerged as a significant information filtering system. It uses several techniques for recommendation which includes content-based, collaborative and hybrid methods. However in spite of all this advancement, the trends followed by recommender system of current generation surveyed in this paper still requires further modifications to make better recommendation. We reviewed various problems that are faced by techniques followed by recommender system and discussed possible solutions that can help in providing better recommendations.

REFERENCES

[1] Francesco Ricci and Lior Rokach and Bracha Shapira, "Introduction to Recommender Systems Handbook ", Springer, 2011

[2] How Computers Know What We Want — Before We Do (http://www.time.com/time/magazine/article/0,9171,1992403,00.html)

[3] J. B. Schafer, J. Konstan, and J. Riedl, "Electronic commerce recommender applications", Journal of Data Mining and Knowledge Discovery, Vol.5,2001

[4] Mitchell Kapor, "Mitchell kapor Quotes", (www.brainyquote.com/quotes/quotes/m/mitchellka163583.html"

[5] Davidson, James, Blake Livingston, Dasarathi Sampath, Benjamin Liebald,Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, and Sujoy Gupta. "The YouTube video recommendation system", Proceedings of the fourth ACM conference on Recommender systems - RecSys 10 RecSys 10, 2010

[6] M . Tim Jones, "Introduction to Approaches And Alorithms of Recommender System", 12 December 2013

[7] B.M. Sarwarm et al., "Analysis of Recommendation Algorithms for E-Commerce," ACM Conf. Electronic Commerce, ACM Press, 2000, pp.158-167

[8] RVVSV Prasad and V Valli Kumari, "A Categorial review of recommender systems", 2012

[9] Adamopoulos, Panagiotis, and AlexanderTuzhilin. "On over-specialization andconcentration bias of recommendations :probabilistic neighborhood selection incollaborative filtering systems",Proceedingsof the 8th ACM Conference onRecommender systems - RecSys 14, 2014.

[10] Gediminas Adomavicius. "Context-AwareRecommender Systems" Systems Handbook, 2011, Recommender