International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No. (03), January 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

A Survey on Recommendation System

Tanya Maan[1], Shikha Gupta[2], Dr. Atul Mishra[3]

[1,2]Deptt. of Computer Engineering,
YMCA University of Science and Technology, Haryana (India)
[3]Deptt. of Information Technology & Computer Application,
YMCA University of Science and Technology, Haryana (India)

ABSTRACT

In the past years ,the way of communication have been changed between the web and the users. With the ample amount of data being present on the web,it became very important to use filtering techniques to find our interest of information. Recommendation systems has its applications in almost every field including e-commerce, e-bussiness, e-government, e-library, e-learning etc. In this paper, we introduce various filtering techniques to ease our work of finding relevant information.Content based and collaborative techniques are the most widely used and are discussed in the paper with their pros and cons. However, the newest research in this area- Hybrid recommendations are also included in this paper.

Keywords: Recommendation Engine, Machine learning, Collaborative filtering, Content based filtering, Hybrid filtering.

I. INTRODUCTION

Since the appearance of the first papers on collaborative filtering since the mid-1990s [1,2,3], Recommender systems became an important research area.The interest in this area still remains high because of its applications, which help users deal with the information overload and provide personalized recommendations to them.

With the immense increase in the amount of information available, retrieving information of our interest has became very difficult. With the development of Information Retrieval systems such as Google, Alta Vista, the task of retrieving information has become quite easy but upto some extent.

The Recommendation systems have fully solves this problem. It has now become the most powerful tool in electronic commerce. In recent years, Recommendations systems have become immensely popular and are used in variety of areas such as in movies, in dating websites, in restaurants, in social tags and in news.

Recommendation systems are softwares that provide suggestions for items to be of use to the user.The suggestions can be related to various decision-making processes such as which item to buy, which news to read , which movies to watch , which music to listen. Over the years, various approaches of recommender systems have been created - collaborative ,content-based filtering.However the newest technology "hybrid recommendations-mixture of collaborative and content' also comes into picture now a days.

Recommender system are good for both the users and the service providers[4].They actually reduce cost of finding and then selecting items[4].In e-commerce recommender system is an effective way to increase profits.They are very useful in libraries.

Item is the term which is used to denote what the system recommends to users[7]. In the immortal words of Steve Jobs: "A lot of times, people don't know what they want until you show it to them."

II.RECOMMENDATION FILTERING TECHNIQUES

A recommendation system is used to filter data for the users in order to make this work easy for them [5].

2.1 Content based filtering

This is the most famous filtering and is used widely .Content based recommendation is targeted at personal level and considers individual preferences and contents of the products for generating recommendations.

Basically these content based recommendations system elaborate a specific profile of each content and then perform some correlation matrices on data.Various items are compared with items previously rated by the user and best matching items are recommended. Collaborative algorithm does two main tasks :

- creating a user profile that describes the types of items the user likes/dislikes
- comparing the user profile to some reference characteristics (with the aim to predict whether the user is interested in an unseen item)

Various advantages of content based filtering are :

- (a)Content based systems are user dependent.
- (b) These systems are highly transparent.

Challenges of content based filtering:

- (a)Scalability : Content based filtering is not scalable.

2.2 Collaborative filtering Recommendations

Another filtering technology that is widely used in recommender systems is Collaborative Filtering.As compared to content based filtering ,collaborative recommender system can automatically filter the information that the content based system could not represent and gives up to date recommendations and informations [6].

Collaborative filtering was developed in late 1990's and it is the most famous filtering technique till date .Various online web services such as Amazon and Netfix are utilizing this filtering techniques.

Collaborative filtering algorithms usually separated into two parts:

- Model-based algorithm.
- Memory-based algorithm.

2.2.1 Memory-based Collaborative Filtering

Memory based collaborative filtering store the entire customer ratings into their memory.

They are also called lazy recommendation algorithms. They do not immediately attempt to calculate customer precedence of an object. Typical examples of this approach are neighbourhood-based CF and item-based/user-based top-N recommendations.

Two types of memory based CF are there:

- Item/Object-based filtering.
- User/Customer-based filtering.

Item/Object-based filtering : Calculate similarity between items and make recommendations. Items usually don't change much, so this often can be computed off line. *Item/Object-based filtering* was recommended by Sarwar et al at 2001 [7]. Items/objects are compared for similarity. The neighborhood of most likely objects is recognized, for every object that belongs to the customer who is active .

For Item/object Based [8]

Step 1: For every object "o" that "c" has preference for yet

Step 2: For every object "p" that "c" has a preference for

Step 3: Compute a similarity s between "o" and "p"

Step 4: Add "c"'s preference for "p", weighted by s, to a running average

Step 5: Return the top objects , ranked by weighted average

User/Customer-based filtering*:* Recommend items by finding similar users. This is often harder to scale because of the dynamic nature of users.

For User/customer Based [8]

Step1: For every object "o" that "c" has no preference for yet

Step 2: For every other customer "d" that has a preference for "o"

Step 3: Compute a similarity s between "c" and "d"

Step 4 :Add "d"'s preference for "o", weighted by s, to a running average

Step 5: Return the top objects, ranked by weighted average

2.2.2 Model-based Collaborative Filtering: memory based tackle the task of "guessing" how much a user will like an item that they did not encounter before. For that they utilize several machine learning algorithms to train on the vector of items for a specific user, then they can build a model that can predict the user's rating for a *new* item that has just been added to the system.

Popular model-based techniques are Bayesian Networks, Singular Value Decomposition, and Probabilistic Latent Semantic Analysis[9].

Challenges of collaborative filtering :

- Scalability[6]: There are millions of users and items present.Thus a large amount of computation power is important to calculate recommendation. For example, with millions of customers (C) and millions of distinct items (O), a CF algorithm with the complexity of O(n) is already too large. Also, many systems need to react immediately to online requirements and make recommendations for all users regardless of their purchases and ratings history, which demands a high scalability of a CF system.
- Data sparsity[10] : If a customer or user has evaluated very few items then its quite difficult to know his taste and his preferences and in this case he could be related to wrong neighbourhood[8].So, this lack of information is quite harmful.

- Cold start problem[4] : This is a situation where a recommendor system do not have adequate information about a customer or object in order to make relevant predictions.

2.3 Hybrid recommendations system

Hybrid systems are the newest recommendor system and combines the best feature of both content based and collaborative based filtering .

- Types of hybrid systems are :
- Weighted hybridisation[4] : This technique combine the results of collaborative and content based filtering to generate a prediction by integrating the techniques of both the recommendations. It increases the performance of the whole system.The main idea of using this technique is to conquer the disadvantages of any individual technique. It computes the weighted sum of the scores and then compare them.
- Switching hybridisation[11] : This technique switches to other recommendations techniques based on the needs of the moment.It solves the new user problem .This method solves the problem which is specific to any one method.One example of switching hybrid is "Daily Learner" that switches between content and collaborative filtering according to the need.
- Mixed hybridisation[12] :Instead of having only one recommendation per object, this technique combine results of different recommendations techniques at the same time.Each object have many recommendations associated from different recommendations techniques.In this technique the result of an individual does not affect the performance in general.
- Cascade Hybridisation [13]: Refinement is done in cascade hybridisation. Results from one recommendations acts as input to other recommendation techniques. Due to this, this process is tolerant to noise.Example of Cascade hybridisation is EntreeC that used a collaborative recommender and knowledge based recommender.

III. PHASES OF RECOMMENDATION ENGINE

3.1 Loading and formatting data

Dataset is a collection of interests and likes of various users by using which we can recommends products and items to them..Dataset can be downloaded from various websites such as movielens,frappe, CoMoDa Dataset.

3.1.1 Data collections are done by two methods-explicitly and implicitly.

(a) Explicit data include the following:

- Asking a customer to rate an object.
- Asking a customer to rank a collection of object from most favourite to least favourite.
- Presenting two object to a customer and asking to choose the better one of them.
- Asking a customer to create a list of object that they likes.

(b) Implicit data collection includes the following:

- Keep observing the object that a customer views in an online store.
- Analyzing its viewing time.
- Keeping a track of the object that a customer purchases online.

3.1.2 Datasets are normally in the form [14] :

{ critic,title,rating}

where critic is the user,

title is the item it rated,

rating is the rating given by the user.

The next step is to arrange the data in a format that is useful to build the recommendation engine .The current data contains a row containing critic,title and rating.This has to be converted to matrix format containing critics as rows,title as columns, and ratings as the cell values.

The data can be viewed as :

	User 1	User 2	User 3	User 4
Item 1	2.5		3.5	
Item 2	4	3	1	3
Item 3		1	5	

Fig 1: Data table for items and users.

3.2 Calculating similarity between the users

This is the very important step as we have to find the similarity between two users to recommend items to them. Various similarity measures are available [15].

1. Distances

We often want to compare two feature vectors, to measure how similar they are. We hope that similar patterns will behave in a similar way.

- Hamming distance

2. Similarity using correlation

The distance is not well normalised .So, we use correlations.

- Normalisation
- Pearson Correlation Coefficient

3. Recommending items to users

The final step is to recommend items to the users.

IV.CONCLUSION AND FUTURE WORK

This paper presented the various techniques and algorithm to build the recommender system. We study various research paper and realised that Collaborative filtering is the mostly used filtering technique but there are various problems related to this such as data sparsity,cold start problem ,scalability. We have also introduced various advantages and disadvantages related to these techniques. Various areas where still much research is to be done in coming years has also been discussed (hybrid recommendations).

REFERENCES

[1]. Hill, W., L. Stead, M. Rosenstein, and G. Furnas. Recommending and evaluating choices in a virtual community of use. In Proceedings of CHI'95.

[2]. Resnick, P., N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl. GroupLens: An open architecture for collaborative filtering of netnews. In Proceedings of the 1994 Computer Supported Cooperative Work Conference, 1994.

[3]. Shardanand, U. and P. Maes. Social information filtering: Algorithms for automating 'word of mouth'. In Proc. of the Conf. on Human Factors in Computing Systems, 1995.

[4]. Recommendation systems : principles,methods and evaluations by F.O.Isinkaye,Y.O.Folajimi,B.A.Ojokoh

[5]. Introduction to Recommender Systems Handbook by Francesco Ricci, Lior Rokach and Bracha Shapira

[6]. A Survey on Recommender Systems based on Collaborative Filtering Technique by Atisha sachan.

[7]. Sarwar.B, Karypis.G, Konstan.J, and Riedl.J. (2001). "Item-Based Collaborative Filtering Recommendation Algorithms". In WWW '01: Proceedings of the 10th International Conference on World Wide Web, New York, NY, USA.11.

[8]. Item based and user based recommendation difference in Mahout by data science

[9]. Memory-Based vs. Model-Based Recommendation Systems by yasserebrahim

[10]. Survey Paper on Recommendation System by Mukta kohar,Chhavi Rana

[11]. An Improved Switching Hybrid Recommender System Using Naive Bayes Classifier and Collaborative Filtering by Mustansar Ali Ghazanfar and Adam Pru¨gel-Bennett ∗

[12]. A mixed hybrid recommender system for given names by Rafael Glauber1, Angelo Loula1, and Jo˜ao B. Rocha-Junior2

[13]. A Hybrid Recommender System Using Link Analysis and Genetic Tuning in the Bipartite Network of BoardGameGeek.com by Brett Boge.

[14]. Building recommendation engines by By Suresh Kumar Gorakala

[15]. Similarity and recommender systems by Hiroshi Shimodaira∗

AUTHORS

Tanya is a M.Tech scholar in "YMCA University of Science and Technology",Faridabad.
(tanyamaan28@gmail.com)

Shikha Gupta is an Assistant professor in department of Computer Engineering of YMCA University of Science and Technology. She is M.tech and her research area is Big Data and Online Marketing.
Shikha.0909@gmail.com

Dr. Atul Kumar Mishra is Professor & Chairman in department of Information Technology & Computer Application of the YMCA University of Science and Technology. He has completed Ph.D. (Computer Engineering) in 2012 from MDU, Rohtak and MS (Computer Science & Technology) in 1992 from IIT Roorkee. His area of research includes MANET, Big Data, Cloud Computing and network management. mish.atul@gmail.com