Recommender System: A Review

Shipranjali[1], Sudhir Kumar Sharma[2]
[1,2]Institute of Information Technology and Management Janakpuri, New Delhi, India
[1]shipra08anjali@gmail.com,[2]sharmasudhir08@gmail.com

***Abstract*- There are number of overwhelming choices over the internet therefore there is a need to filter and efficiently deliver the relevant information so that we can overcome the information overload problem, a potential problem of many internet users. Thus recommender systems provide users the solution to this problem. They search through a huge volume of dynamically generated information to provide users with their choice of content, what they want to see. Various techniques have been proposed and many software have been developed for a different no. of applications. This paper therefore reviews and explore up to date applications of recommender system, different characteristics and types of prediction techniques in recommendation systems.**

***Keywords*- Recommender Systems, Online overload, Application development**

I. INTRODUCTION

During the last few decades, with the growth in no. of available online information, e commerce websites and the visitors to internet have created a potential challenge of information overload. The search engines or information retrieval systems have partially solved the problem but they lack in personalization and prioritization. Therefore recommender systems were increased which fulfill this missing feature. We are aware that nowadays, people are moving further on buying products online instead of going to the market for the product. Also sometimes when the products they want are recommended to user it boost up their confidence for buying the same. Therefore we can say that recommender systems are the algorithms which are targeted at suggesting relevant items to users (items can be any text to read, any movie to watch, any products to buy, etc.). Recommendation systems can be beneficial in various ways like they can help in driving traffic, delivering relevant content, reduce workload and overhead, creating customer satisfaction, increasing revenues, etc. Therefore they are beneficial to both service provider and users. Generally, a system which will be providing fast and accurate recommendations will be able to attract the customers' interest and will bring profit to companies. A recommender system usually use one of the three mentioned ways to produce a list of recommendations: Collaborative filtering (CF), Content-based filtering, and Hybrid recommender systems.

In this paper, the discussion is about the two major conventional recommendation techniques and highlighted their strengths and challenges, comparison between the two, also the hybridization technique which overcome the shortcomings of other two techniques. Lastly we discusses the applications of recommender system.

The paper in the first section discusses about the recommendation technique, after that second section discusses about content-based filtering including the pros, cons and example of the same. The third section discusses about the collaborative filtering again including type, pros, cons and example. Fourth section discusses about the comparison of two techniques mentioned above. Next section discuss about the hybrid filtering, about their methods and the second last section discusses about the applications of recommender system. The conclusion is presented in the last section.

II. RECOMMENDATION TECHNIQUE

The use of productive and appropriate technique is necessary so that the system will provide accurate and appropriate recommendation to its individual user. Fig 1 demonstrates the types of recommendation filtering techniques.

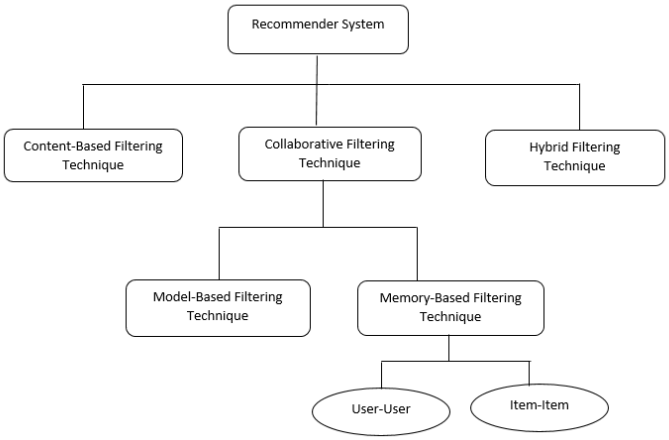

Fig 1: Types of recommendation system techniques

III. CONTENT-BASED FILTERING

This technique is based on the information or description provided for the particular product. Based on the context or description of the item the system will find the similarity between items. Also for finding similar products which user may like the user's previous history is taken into account. For example, there is a user one who liked a Marvel movie then we can recommend him superheroes

movies or also the movies which are having the similar actors as in Marvel movie. In this technique two type of data is used, first being the interest of user, whatever the user like, second being the user information, their personal information like age and sometimes their history too. Text document plays mostly a major role of information source in this type of systems. A standard approach for term parsing selects single words from documents. The vector space model and latent semantic indexing are two methods that use these terms to represent documents as vectors in a multi-dimensional space [4].

This type of filtering technique uses properties or characteristics of an item to provide recommendations. Characteristic information includes:

- Characteristics of Items (Keywords and Attributes)
- Characteristics of Users (Profile Information)

Fig 2 shows how content based filtering works. It is showing that a similar movie to the movie watched earlier by the user can be recommended to him/her.

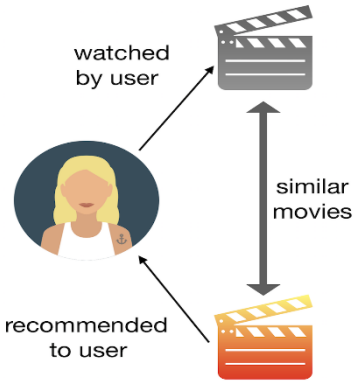

Fig 2: Working of content based filtering

A. *Pros and Cons of Content-Based Filtering*

The following are the pros of content-based filtering:

- User independence:

In this technique there is no dependency on other users for providing a recommendation to a user. As we only have to analyze the user profile and items for recommending any item.

- Transparency:

In this technique the user can know on what basis or features an item is recommended to them. Instead of collaborative filtering where you are getting recommendations because you have the similar taste as of any other unknown user.

- No Cold start problem:

Here in this method a new item can be easily recommended to a user instead of waiting for the raring of some other users.

The following are the cons of content-based filtering:

- Limited content analysis:

If there is no enough information present to discriminate the items precisely, then the recommendation will also be not precise in the end.

- New user:

Let say if there is a new user in the system and there is not enough information available to build its profile then the recommendation could not be accurate.

- Serendipity Problem:

This technique do not have inherent method for finding something unexpected. It only suggests items whose rating are high when matched against user profile, hence user will be recommended items similar to those which they have already rated.

B. *Example of Content-Based Filtering*

Following is taken an example to understand how this technique works in recommending a movie to a Netflix user named Joy as per Fig 3 (a).
Let's assume that Joy gave good rating to movies like "Avengers" and "Captain America" which are marked as "Fiction" genre and gave a bad rating to the movie "Titanic" which is marked as "Romantic" genre.

Joy

Movies	Reviews Given	Ratings
Avengers	Yes	Good
Captain America	Yes	Good
Titanic	Yes	Bad

Fig 3: (a) Example of content based filtering

Now we had created a user vector for Joy based on his given three ratings as per Fig 3 (b):

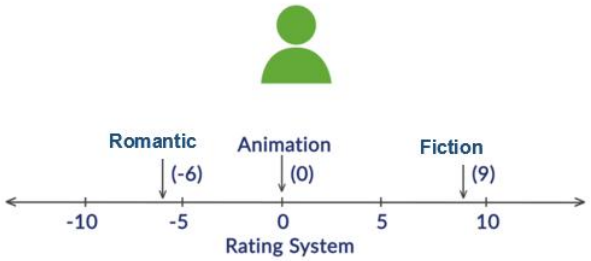

Fig 3: (b) Example of content based filtering

Taking a rating scale of -10 to 10, since Joy like Fiction movies, we give value of 9 to "Fiction", Joy hasn't seen any Animation movies, we give 0 to "Animation" and since Joy has disliked movies with Romantic genre, we give '-6 ' to "Romantic".
Therefore user Vector for Joy will be (9, 0, -6) in order of (Fiction, Animation, Romantic) as per Fig 3 (c).

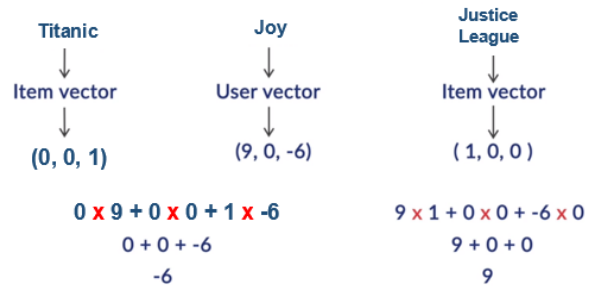

Fig 3: (c) Example of content based filtering

The item vector for movie "Titanic" is (0, 0, 1) and the movie "Justice League" is (1,0,0) in order of (Fiction, Animation, Romantic).
Now we need to take dot product of two 2-D vectors: Item vector and User Vector as per Fig 3 (d).

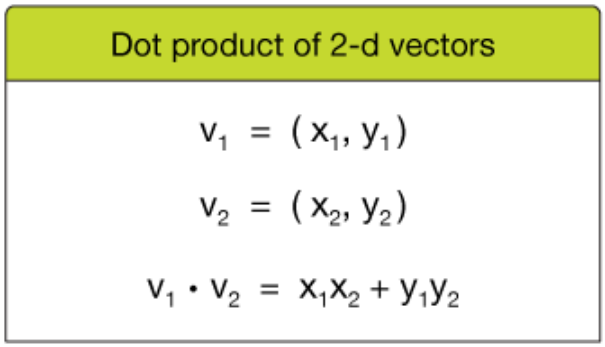

Fig 3: d) Example of content based filtering

Accordingly, the dot product of "Titanic" is -6 and that of "Justice League" is 9.
Hence "Justice League" will be recommended to Joy, which also match our instinct that Joy likes Fiction movies and dislikes Romantic movies.
Similarly, we can calculate the dot products of all the item vectors of all the movies present and recommend top 10 movies to Joy.

So this is how content-based filtering technique works.

News Dude, a personal news system which works by utilizing speech to read news stories to users. For describing news stories so that short-term recommendations can be generated it uses TF-IDF (Term Frequency- Inverse Document Frequency) model which is further compared with Cosine Similarity Measure and lastly supplied to a learning algorithm.

IV. COLLABORATIVE FILTERING

Collaborative Filtering technique is based on the past interactions between user and items in order to produce new recommendation, and these interactions are kept in "user-item interactions matrix". For Example a user "A" rated movie m1, m2, m3 with good rating and another user "B" rated movie m1, m3, m4 with good rating. Therefore now we can recommend movie m2 to user "B" and m4 to user "A" because they have similar taste in movies.
The main idea behind collaborative filtering is these previous user-item interactions are enough to trace similar users or similar items and further make predictions depending on these estimated closeness results.
The collaborative technique is further divided into 2 sub categories:
First memory based and second model based approaches. In memory based technique it is based on nearest neighbor search, the recorded interactions, that is search for the users which are closest among a user of interest and then recommend the user item which is most popular among their neighbor. In model based technique assume an underlying "productive" model which explains the user-item interactions and try to discover it in order to make new predictions.
The Fig 4 explains the working of collaborative filtering. It is showing that a user is having similar taste (pizza, pasta) like another user therefore a third item liked by second user (cold drink) can be recommended to first user.

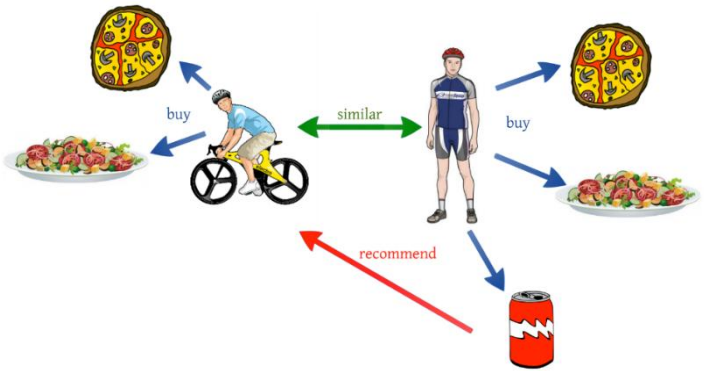

Fig 4: Working of Collaborative Filtering

A. *Types of collaborative Filtering*

- Memory based:

For computing similarity between users or items it uses user rating data. This is used for making recommendations. It is an earlier mechanism and was used in many

commercial systems. It is effective and easy to implement. Neighborhood based CF and item/user-based top N recommendations are some typical examples of this mechanism. The Representative techniques are Neighbor-based CF (item-based/user-based CF algorithms with Pearson/vector cosine correlation) and Item-based/user-based top-N recommendations [2].

User-user:
This method is said to be "user-centered" as it represent users based on their interactions with items and evaluate distances between users. This method tries to identify users with similar interaction profiles so that most popular item could be suggested among these neighbors.
Let's assume we want to make a recommendation for a given user. Firstly, each user can be represented by their vector of interactions with the different items as shown in Fig 5. Then we can find some type of similarity among those users for which we are going to make recommendation and every other user. That similarity will be measured such that if there are two users having similar interaction on similar item then they will be considered as being close. Once the similarities to every user are found, then we can keep the k-nearest neighbors to our user and then further most popular items can be suggested among them.

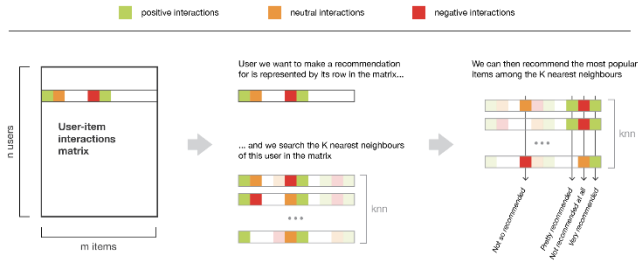

Fig 5: Illustration of the user-user method.

Item-Item:
This method is said to be "item-centered" as it represent items depending upon the interactions user had with them and calculate the distances between those items. In this method we try to find items similar to those with which the user has positively interacted with and if many user has interacted with both the items in a similar way then these two items will be considered similar.
Let's assume that we want to make a recommendation for a given user. So, firstly, we will take the item this user liked the most and represent it by its vector of interaction with every users as shown in Fig 6. Now we can find the resemblance between the "best item" and the remaining items. When we found all the resemblances, we will keep the k-nearest neighbors to the selected item that are new to our user of interest and recommend these items.

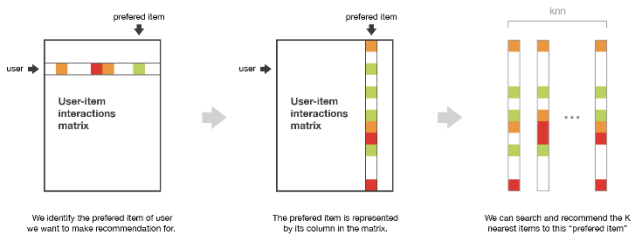

Fig 6: Illustration of the item-item method.

- Model based:

Methods based on this type assume some latent interaction model. Here the model is trained such that from its own representation of users and items it rebuild user-item interactions values. Further based on this model new suggestions can be done. The latent representation of users and items extracted by this model contain a mathematical meaning which can be tough to interpret for a human being.
One of the algorithms used for model based approach is matrix factorization.
Let us understand with the help of an example how it works.
Let's consider we are having a matrix of user-movie rating. Now for modeling the interactions between users and movies, we can assume that:

1. There will exist some features describing movies pretty well.
2. These existing features can also be utilized for describing user preferences like more values for features user likes and less values otherwise.

Though we do not want to give explicitly these features to our model. Instead we want to let the system find these useful features for making its own representations of both users and items by itself.
As they are learned and not given, extracted features taken individually have a mathematical meaning but no intuitive interpretation (and, so, are difficult, if not impossible, to understand as human) [1].
As here Fig 7 is illustrating the matrix factorization method where the user-item interactions matrix is assumed to be equal to the dot product of user matrix and a transposed item matrix.

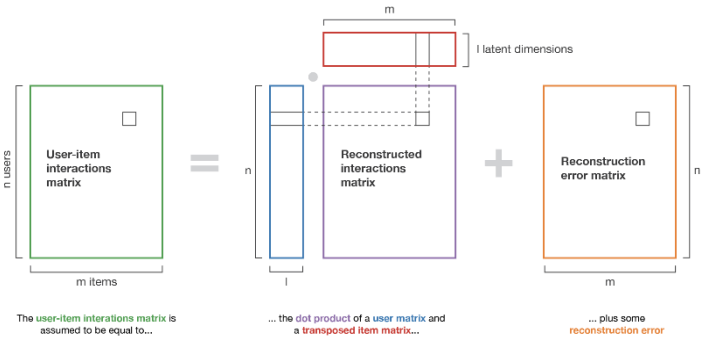

Fig 7: Illustration of the matrix factorization method.

B. *Pros and Cons of Collaborative Filtering*

The following are the pros of content-based filtering:

- No domain knowledge necessary
  Here we are not requiring domain knowledge because the embedding are automatically learned.

- Serendipity
  With the help of this technique users are able to discover new interest. Because of similar user interest an item will be recommended to user whether the user is interested or not.

The following are the cons of content-based filtering:

- Cold-start problem
  This is the situation when the system lacks in information about user or item to make any relevant predictions. The system will be having problem in knowing the taste of new user because they have not rated any item.

- Data sparsity problem
  This problem arise when there is lack of information, means that when only some of the total number of items which are available in a database are rated by users.

- Synonym
  This is the problem of very similar items to have different names or entries. It is seen that many recommender systems finds it difficult to create difference among closely related items e.g. baby wear and baby cloth.

C. *Examples of Collaborative-Based Filtering*

BestBuy, an American company uses recommender system to improve their customer experience. The company's strategy is based on query search and click data. Since 2015, Best Buy has used the information for trying to predict what customers are interested in. In this for making customer recommendations the query-based and item-to-item system make cluster models which allow them for making recommendations.
Also Amazon too uses a mix of user-user and item-item collaborative filtering as recommendation technique.
Ringo is a collaborative filtering (user-based) system built for making recommendations of music albums and artists. In this, initially when user enters the system, they are provided a list of about 125 artists to rate them according to their interest that is how much he/she likes listening to them.

V. COMPARISON BETWEEN COLLABORATIVE AND CONTENT BASED RECOMMENDER SYSTEM

TABLE I
COMPARISON

Content-based filtering	**Collaborative filtering**
Based on the context or description of the item the system will find the similarity between items.	Based on the past interactions between user and items in order to produce new recommendation
Uses properties or characteristics of an item to provide recommendations.	Uses previous user- item interactions to trace similar users or items.
Suffer no cold start problem.	Suffer cold start problem.
It only suggests items whose rating are high when matched against user profile hence lack in discovering new interest.	With the help of this technique users are able to discover new interest.
It is not further classified.	It is further classified in memory based and model based techniques.
Genetic algorithms, neural networks and the Bayesian classifier are among the algorithms used in it for learning techniques.	Nearest neighborhood algorithm is the standard algorithm for this technique.

VI. HYBRID FILTERING

We have discussed above the two major techniques of recommender systems, also we saw their pros and cons. Therefore a techniques was formed which combines different recommendation techniques for obtaining better results known as hybrid filtering technique. The main idea behind this technique was that if we will merge two algorithms or techniques then we can have more accurate and effective results also the disadvantage of one can be overcome by the other. The combinations can be done in any of the ways: content via collaboration, collaboration via content, separate implementation and combining the results or building a system that brings together both approaches.

TABLE II
HYBRIDIZATION METHODS

Hybridization method	**Description**
Weighted	In this method the votes (or scores) are combined together taken from several

	recommendation techniques so that a single recommendation can be produced [7].
Switching	In this method depending upon the current situation the system switches between techniques [7].
Mixed	In this method recommendations taken from several different recommenders are presented at the same time [7].
Feature combination	In this method features taken from different recommendation data sources are put together in a single algorithm [7].
Cascade	In this method recommendations given by other system is refined by one system [7].
Feature augmentation	In this method the outcome of one technique works as input feature to another [7].
Meta-level	In this method model learned by one system works as input to another [7].

VII. APPLICATIONS OF RECOMMENDER SYSTEM

The applications of recommender systems include recommending movies, websites, books, documents, television programs, tourism scenic spots, music, conferences and learning materials, and involve the areas of e-commerce, e-library, e-business services, e-government and e-learning [9].
The most common recommender systems applications include:

- E-commerce: recommendations for customers while buying online products like gadgets, books, clothes, etc.
- Entertainment: recommendations for movies and music.
- Content: e-learning, recommendation for web pages, e-mail filters, personalized newspapers, etc.
- Services: matchmaking services, recommendation of experts for consultation, - recommendations of travel services, recommendation of houses to rent.

VIII. CONCLUSION

Recommendation systems are everywhere. They help us in gaining new opportunities for getting personalized information on the internet. It is expected from a good recommendation system that time to time it will be able to give positive and appropriate recommendations and also give alternative recommendations so that the fatigue of the users can be broken built by seeing the same recommendation list again and again. In this paper the discussion was about the two major conventional recommendation techniques and highlighted their strengths and challenges, also the hybridization technique which overcome the shortcomings of other two techniques. Lastly we shortly discussed the applications of recommender system.

REFERENCES

[1] Website- *https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada*
[2] P. N. Vijaya Kumar *A Survey on Recommender Systems (RSS) and Its Applications (2014)*
[3] Website- *https://towardsdatascience.com/brief-on-recommender-systems-b86a1068a4dd*
[4] Website- *http://recommender-systems.org/hybrid-recommender-systems/*
[5] Website- *https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada*
[6] Website- *https://www.sciencedirect.com/science/article/pii/S1110866515000341#b0005*
[7] Robin Burke *Hybrid Recommender Systems: Survey and Experiments*
[8] Dietmar jannach, Markus zanker *Recommender Systems An Introduction*
[9] Jie Lu *Recommender system application developments: A survey (2015)*
[10] Website- *https://www.iteratorshq.com/blog/an-introduction-recommender-systems-9-easy-examples/*