RESEARCH REPORT
ON
FILTERING TECHNIQUES

Submitted By:

SUPARNO BANERJEE

Data Science Batch 2

MIT WPU
Mail: 2021031@risewpu.com

1. INTRODUCTION

Nowadays, people tend to buy products online more than from stores. Earlier, people used to buy products based on the reviews given by relatives or friends but now as the options increased and we can buy anything digitally we need to assure people that the product is good, and they will like it. To give confidence in buying the products, recommender systems were built.

Recommender systems are information filtering systems, that deal with the problem of information overload, by filtering vital information out of a large amount of dynamically generated information according to user's preferences, interest, or observed behaviour about item. Recommender system could predict whether a particular user would prefer an item or not based on the user's profile.

Recommender systems are beneficial to both service providers and users. They reduce transaction costs of finding and selecting items in an online shopping environment. Recommendation systems have also proved to improve decision making process and quality. e.g. In e-commerce setting, recommender systems enhance revenues, for the fact that they are effective means of selling more products. Therefore, the need to use efficient and accurate recommendation techniques within a system that will provide relevant and dependable recommendations for users cannot be over-emphasized.

2. RECOMMMENDATION ENGINE

Recommendation engines filter out the products that a particular customer would be interested in or would buy based on his or her previous buying history. The more data available about a customer the more accurate the recommendations.

But if the customer is new, this method will fail as we have no previous data for that customer. So, to tackle this issue different methods are used; for example, often the most popular products are recommended. These recommendations would not be must accurate as they are not customer dependent and are the same for all new customers. Some businesses ask new customers their interests so that they can recommend more precisely.

2.1 PHASES OF RECOMMENDATION PROCESS

2.1.1 Information Collection Phase

This phase collects relevant information of users to generate a user profile or model for the prediction tasks, including user's attribute, behaviour or content of the resources the user accesses. A recommendation agent cannot function accurately until the user profile/model has been well constructed. The system needs to know as much as possible from the user to provide reasonable recommendation right from the onset.

2.1.2 Learning Phase

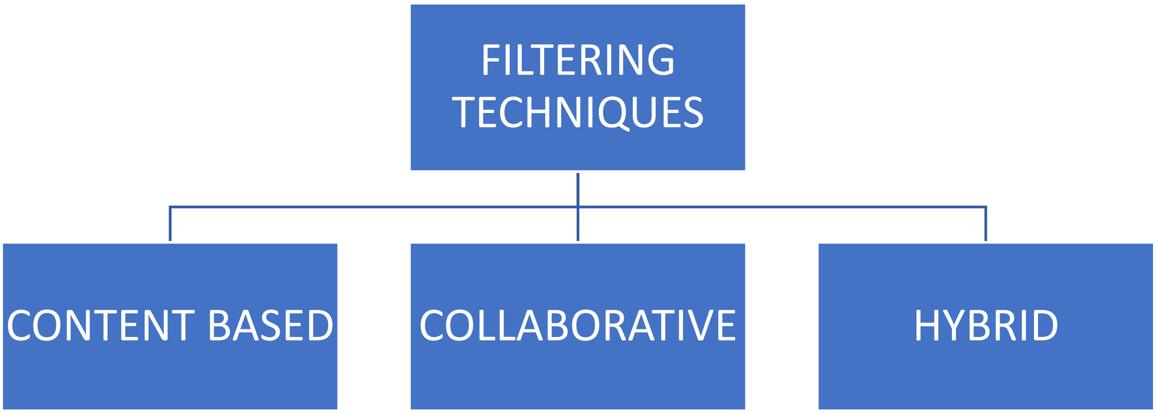
It applies a learning algorithm to filter and exploit the user's features from the feedback gathered in information collection phase.

2.1.3 Recommendation Phase

It recommends or predicts what kind of items the user may prefer. This can be made either directly based on the dataset collected in information collection phase which could be memory based or model based or through the system's observed activities of the user.

Recently, various approaches for building recommendation systems have been developed, which can utilize either collaborative filtering, content-based filtering or hybrid filtering.

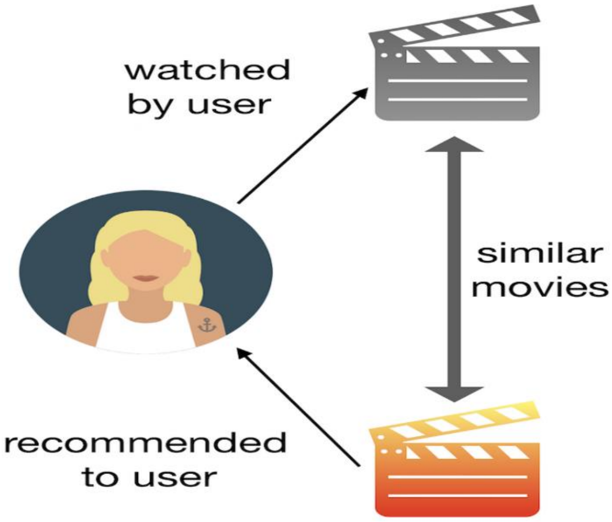
3. FILTERING TECHNIQUES



3.1 CONTENT BASED FILTERING

This filtering is based on the description, or some data provided for that product. The system finds the similarity between products based on its context or description. The user's previous history is considered to find similar products the user may like.

For example, if a user likes movies such as 'Fast & Furious' then we can recommend him the movies with tags like - 'Vin Diesel', 'Action', 'Cars' etc.



In this filtering, two types of data are used –

First, the likes of the user, the user's interest, user's personal information such as age or, sometimes the user's history too. This data is represented by the user vector.

Second, information related to the product known as an item vector. The item vector contains the features of all items based on which similarity between them can be calculated.

The recommendations are calculated using cosine similarity. If 'A' is the user vector and 'B' is the item vector then cosine similarity is given by:

$$cos(\theta) = \frac{A \cdot B}{\|A\ \|\ B\|} = \frac{\sum_i A_i B_i}{\sqrt{\sum_i A_i^2}\sqrt{\sum_i B_i^2}}$$

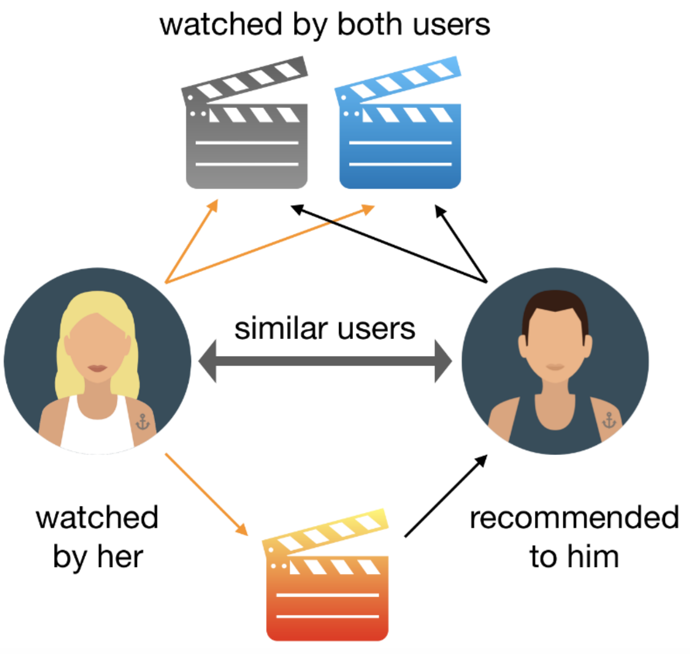
Values calculated in the cosine similarity matrix are sorted in descending order and the items at the top for that user are recommended.

PROS	CONS
• Has the ability to recommend new items even if the database does not contain user preferences. • If the user preferences change, it has the capacity to adjust it's recommendations in a short span of time.	• Recommendation is dependent on the item's metadata. Hence, require rich description of items and very well organised User profile. • Users are restricted to getting recommendations similar to the items already defined in their profiles

3.2 COLLABORATIVE FILTERING

This technique works by building a database (user-item matrix) of preferences for items by users. It then matches users with relevant interest and preferences by calculating similarities between their profiles to make recommendations. Such users build a group called neighbourhood. A user gets recommendations to those items that he has not rated before but that were already positively rated by users in his neighbourhood.

For example, if the user 'A' likes 'Avengers', 'Spider Man' and 'Venom' while the user 'B' likes 'Avengers', 'Spider Man' and 'Captain America' then they have similar interests. So, there is a huge probability that the user 'A' would like 'Captain America' and the user 'B' would like 'Venom'.



The technique of collaborative filtering can be divided into two categories: memory-based and model-based.

3.2.1 MEMORY BASED

The items that were already rated by the user before, play a relevant role in searching for a neighbour that shares appreciation with him. Once a neighbour of a user is found, different algorithms can be used to combine the preferences of neighbours to generate recommendations. It can be achieved in 2 ways:

a. **User based collaborative technique**: the user vector includes all the items purchased by the user and rating given for each product. The similarity is calculated between users using an n*n matrix in which n is the number of users present. The similarity is calculated using the same cosine similarity formula. Now, the recommending matrix is calculated. In this, the rating is multiplied by the similarity between the users who have bought this item and the user to which item has to be recommended. This value is calculated for all items that are new for that user and are sorted in descending order. Then the top items are recommended to that user.
b. **Item based collaborative technique**: in this, rather than considering similar users, similar items are considered. If the user ‘A’ loves ‘Inception’ he may like ‘The Martian’ as the lead actor is similar. Here, the recommendation matrix is m*m matrix where m is the number of items present.

3.2.2 MODEL BASED

This technique employs the previous ratings to learn a model to improve the performance of Collaborative filtering Technique. The model building process can be done using machine learning or data mining techniques. It analyses the user-item matrix to identify relations between items; they use these relations to compare the list of top-N recommendations. Examples of these techniques: Singular Value Decomposition (SVD), Matrix Completion Technique, Latent Semantic methods, and Regression and Clustering etc.

PROS	CONS
• The larger the data set (the more the business is scalable), more satisfactory recommendations. • it can recommend items that are relevant to the user even without the content being in the user’s profile.	• User’s previous history is required or data for products is required • If there are few items rated by the user, then it leads to sparse user-item matrix and hence, unable to locate successful neighbours and thus, generation of weak recommendations.

3.3 HYBRID FILTERING

Hybrid filtering technique combines different recommendation techniques in order to gain better system optimization to avoid some limitations and problems of pure recommendation systems. The idea behind hybrid techniques is that a combination of algorithms will provide more accurate and effective recommendations than a single algorithm as the disadvantages of one algorithm can be overcome by another algorithm.

Using multiple recommendation techniques can suppress the weaknesses of an individual technique in a combined model. The combination of approaches can be done in any of the following ways:

a. separate implementation of algorithms and combining the result
b. utilizing some content-based filtering in collaborative approach
c. utilizing some collaborative filtering in content-based approach
d. creating a unified recommendation system that brings together both approaches