Approach Towards Hybrid Recommendation System using Content Based and Collaborative Filtering Techniques

Siddhesh Jagtap[1] , Yash Mane[2] , Tushar Kadam[3], Trupti Dange[4]

[1]Siddhesh Jagtap ,Dept of Computer Engineering, RMDSSOE ,India
[2]Yash Mane, Dept of Computer Engineering, RMDSSOE, India
[3]Tushar Kadam ,Dept of Computer Engineering, RMDSSOE ,India
[4]Prof.Trupti Dange ,Dept of Computer Engineering, RMDSSOE ,India

---***---

Abstract – *In today's world, every customer is faced with multiple choices. For example, If a person is looking for food to eat online without any specific idea of what he/she wants, there's a wide range of possibilities and mouth-watering results how the search might pan out. The person might waste a lot of time browsing around on the internet and trawling through various sites hoping to strike gold. Then he/she might look for recommendations from other people.*

In this paper, we describe various techniques and approaches used by the recommender system and the hybrid based recommendation system related research.

Key Words: Hybrid recommendation system, Machine Learning, Content Based, Collaborative filtering, etc

1. INTRODUCTION

In today's world where internet plays a vital role in human life, the users are facing problems of choosing due to the wide range of collection. Searching from a motel to good investment options, there is too much information/data available over the internet. To help the users cope with this information explosion, companies have deployed recommendation systems for guiding their users. The research in this area of recommendation systems has been going on for quite long time but the interest still remains high because of the abundance of practical applications and the problem rich domain.

Recommender systems are used for providing personalized recommendations based on the user profile and previous behaviour. Recommender systems such as Amazon, Netflix, and Youtube are widely used in the internet industry. Recommendation systems help the users to find and select items (e.g. Books, Movies, Food, Restaurants.) from wide collection available one the web or in other electronic information sources. Among a large set of items and a description of the user's needs, they present to the user a small set of the items that are well suited to them. The main purpose of a recommendation system is to give a satisfying personalization depending upon their needs[4].

The process of recommendation system is not so difficult as it uses different techniques like Machine Learning. The techniques used under Machine Learning are Content Based and Collaborative Filtering. As they both have their unique approach towards recommendation system. They also have some problems like cold start, sparsity, etc.

But combining content based and collaborative filtering will result in high accuracy hybrid recommendation system.

2. KEY TECHNOLOGIES

A good recommendation system uses some techniques that gives it a better result. One of the techniques used by recommendation technique is Collaborative Filtering and the other is Content Based.

2.1. COLLABORATIVE FILTERING

Collaborative filtering uses *similarities* between users and items simultaneously to provide recommendations. This allows for serendipitous recommendations; that is, collaborative filtering models can recommend an item to user A based on the interests of a similar user B[6].

Furthermore, the embeddings can be learned automatically without relying on hand engineering of features.

Consider a food recommendation system in which the training data consists of a feedback matrix in which:

- Each row represents a user
- Each column represents an item(Food varities)

Table -1: Food based on user ratings

Illustration of Example			
Menu	**Rating**	**Description**	
Dal Makhani	PG-13	Indian recipe(Veg)	
Paneer Masala	PG	Indian recipe(Veg)	
Chicken Masala	PG	Non-veg Food recipe	
Butter Chicken	R	Non-veg Food	

		recipe	

The feedback about food falls into one of two categories :

- Explicit: Users specify how much they liked a particular food by providing a numerical rating.
- Implicit: If a user order's food from the menu, the system infers that the user is interested/liked[6].

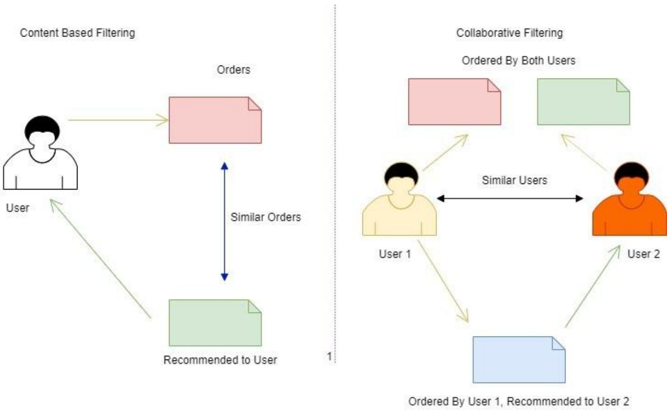

Figure 1: Content based filtering & Collaborative filtering

To simplify, we will assume that the feedback matrix is binary, that is, a value of 1 indicates interest in the particular food item.

When a user visits the homepage, the system should recommend food items from that particular menu based on both:

- Similarity to food items the user has liked in the past
- Food items that similar users liked.

2.2. CONTENT BASED FILTERING

Content here refers to the content or attributes or item set of the products the users likes. So, the idea in content-based filtering is to tag products using certain keywords, understand what the user likes, looks up those keywords in the database and recommend different products with the same attributes[6].

2 different approaches of content based recommendation system:

- Approach 1: Analyzing Description of the Content Only.
- Approach 2: Building User Profile and Item Profile from User Rated Content.

A content based recommender works with data that the user provides either explicitly (rating) or implicitly (clicking on the link). Based on that data, a user profile is generated which is then used to make suggestion to the user. As the user provides more inputs or takes actions on the recommendations, the engine becomes more and more accurate.

2.3. CONCEPTS USED IN CONTENT BASED RECOMMENDER SYSTEM

The concepts of Term Frequency (**TF**) and Inverse Document Frequency (**IDF**) are used in Information retrieval systems and also content based filtering mechanisms (such as a content based recommender). They are used to determine the relative importance of a document/ article/ news items/ movies/ food items, etc.

2.3.1. TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY

TF is simply the frequency of a word in a document. IDF is the inverse of the document frequency among whole corpus of documents. TF-IDF is used mainly because of two reasons: Suppose we search for "***the rise of analytics***" on Google, it is certain that "***the***" will occur more frequently than "***analytics***" but the relative importance of analytic is higher than the search query point of view. In such cases, TF-IDF weighting negates the effect of high frequency words in determining the importance of an item (document).
General Formula for calculating TF and IDF are given below:

$$\mathrm{Tf(t)} = \frac{\text{Frequency occerence of term t in document}}{\text{Total number of terms in documents}}$$

$$\mathrm{Idf(t)} = \log_{10}\left(\frac{\text{Total Number of documents}}{\text{Number of documents containing term t}}\right)$$

2.3.2. COSINE SIMILARITY

We use cosine similarity to quantify the similarities between movies, food items, etc. Cosine similarity ranges from -1 to 1 and is calculated as the dot product between two vectors divided by their magnitudes.

$$\mathrm{sim(A,B)} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

3. HYBRID COLLABORATIVE FILTERING TECHNIQUE

Even though content- based and collaborative filtering are the most popular in the practical applications, both of them suffer from several limitations or we can say disadvantages.

Hybrid recommendation approaches, as a combination of two or more approaches, have been proposed to overcome the disadvantages or limitations of traditional approaches or techniques and improves the accuracy and quality of the recommendation offered to the user by the particular system[1]. A hybrid collaborative recommendation is also known as Content Boosted Collaborative Filtering, methodology is proposed to deal with a huge amount of data intended for careful product grouping to address the data sparsity issue of collaborative filtering and content based filtering[1]. The more the data is generated in the system, more is the accuracy of the recommender system, that is if the user visits a specific site for say food items, the more orders generated, the more specific is the accurate recommendation when he/she visits it next time[9].

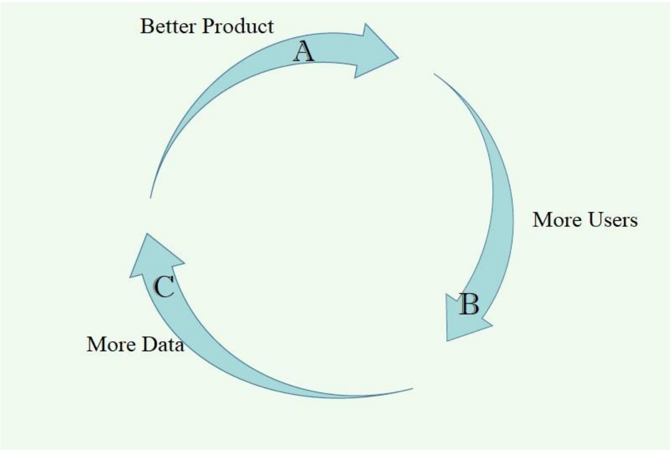

Figure 2: Business cycle for Hybrid recommendation system

The technical scope of collaborative filtering recommendation is way wider than other recommendation techniques. The accuracy of recommendation systems is relatively high but it lags in some ways like cold start problem, sparsity. Same goes with content based recommendation system but it does not rely on user evaluation information so there is limitation like sparsity. Rather it relys on previous data. So, for overcoming these shortcomings, the hybrid recommendation can adopt different weight strategies. So when recommendation in food items is increased, user's evaluation information is more, weight based on these techniques is relatively higher which will result in a good recommendation system.
This knowledge increase makes it especially promising to explore new ways to extent underlying collaborative filtering algorithms with content data and content based algorithm with the user behavior data.

Step 1: Use content-based predictor to calculate the pseudo user-rating vector 'Y' for every user 'X'.

Step-2: Weight all users with respect to similarity with the active user. Similarly between users is measured as the Pearson correlation between their ratings vectors.

Step-3: Select n users that have the highest similarity with the active user. These users form the neighbourhood.

Step-4: Compute a prediction from a weighted combination of the selected neighbors ratings.

General Formula for calculating Pearson Correlation is given below:

$$r=\frac{\Sigma\,(X_i-\overline{X})\,(Y_i-\bar{y})}{\sqrt{\Sigma\,(X_i-\overline{X})^2\Sigma(Y_i-\bar{y})}\ \ ^2}$$

Technically, the Pearson product-moment correlation coefficient (**PPMCC**) is a measure of the direct connection between two variables X and Y, giving a worth amongst +1 and -1 comprehensive, where 1 is all out positive correlation, 0 is no correlation, and -1 is complete negative correlation[5].

Below are the readings taken during food items recommender system using hybrid approach in which correlation is calculated for each item.

Out[136]:

	name	wcorelation
9	Ven Pongal	2.915541
19	Dum Olav	2.915541
16	Dum Olav	2.915541
44	Jhal Muri	2.915541
23	Arhar Ki Dal	2.915541
46	Chepa Pulusu	2.915541
12	Yogurt Lamb Curry	2.915541
40	Fafda	2.915541
10	Yogurt Lamb Curry	2.915541
20	Arhar Ki Dal	2.915541

Figure 3: Calculated correlation for food items.

4. CONCLUSION

In the above paper, we have studied the techniques and approaches related to a good recommender system and how hybrid recommender system overcomes limitations faced by collaborative filtering and content based recommender system togetherly and resulting in a good accuracy in recommending items.

REFERENCES

[1] Geetha G, Safa M, Fancy C , Saranya D "A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System", NCMTA 2018.

[2] Wu Dan,"Music Personalized Recommendation System based on Hybrid filtering", ICITBS 2017

[3] Wang Wenzhen,"Personalized Music Recommendation algorithm based on hybrid collaborative filtering technology", ICSGEA 2017

[4] Ms. Shakila Shaikh, Dr. Sheetal Rathi, Asst Prof. Prachi Janrao," Recommendation system in E-commerce websites: A Graph Based Approached", IEEE 7th International Advance Computing Conference, 2017

[5] Arjun Singh Tomar , Amit Srivastava, Shishir Kumar,"AN IMPLEMENTATION OF PEARSON CORRELATION METHOD FOR PREDICTING ITEMS TO USER IN E-COMMERCE", IJESRT 2016

[6] Shah, K., Salunke, A., Dongare, S., & Antala, K. (2017),"Recommender systems: An overview of different approaches to recommendations", 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).2017

[7] Yannan Song, Wei Ji, Shi Liu, "Research on Personalized Hybrid Recommendation System", IEEE Conference 2017

[8] Bekir Berker Türker ; Resul Tugay ; İpek Kızıl ; Şule Öğüdücü,"Hotel Recommendation System Based on User Profiles and Collaborative Filtering", IEEE 4th International Conference,2019

[9] Karan Dayma, Rohan Parabh , Swapnali Kurhade," A Hybrid Model for Book Recommendation", IEEE 2018 2nd International Conference.