RECOMMENDATION SYSTEM: CHALLENGES, TYPES AND BENEFITS

[1]Archana N L, [2]Dr. Hema Jagadish

[1]M Tech Student, [2]Assistant Professor
[1]Information Science and Technology,
[1]Bangalore Institute of Technology, Bangalore, India

Abstract : Now a days, people are all attached to the virtual world that exists over the internet. Overtime, the main problem users have been facing, ranges from lack of information to information being overloaded. It was never easy to obtain information. But, there is so much information out there that it has been easy to get overwhelmed. It is very important to have some means for filtering out the noise. Recommender systems solves this problem as they are smart filters that would help the users discover what they like. For business, it is the system that increases revenue and drives content. This paper explores the various benefits, challenges and different types of algorithms used for recommendations with a case study for content based and collaborative filtering recommender systems**.**

IndexTerms **— Filtering, Recommendation System(RS), algorithms, Content based and Collaborative filtering.**

I. INTRODUCTION

With the growing amount of information over the Internet and with the significant rise in the number of users, Many opportunities such as sharing information, knowledge and opinions with other users favored the development of social networks such as Facebook. Nowadays, musicians can get famous faster than ever before by uploading their tracks, authors can share their articles with thousands of readers all over the globe. Many online shopping sites have been open over the Internet. Any user from any part of the world can purchase any commodity from any part of the world. The amount of information and goods got extremely huge as there were no limitations for the items over the Internet. This led to information overload. Search engines solved this problem partially. So Recommendation systems were used to solve the problem.

Recommender systems are a software used for filtering and sorting the items and information. It gradually becomes important for any organizations to search, map and provide users with the relevant information based on their tastes and preferences. Organizations, now a day is building smart and intelligent recommendation systems by analyzing the past behavior of the users. Hence providing choices in terms of “Relevant Job postings”, “Movies of Interest”, “Suggested Videos”, “Facebook friends that you may know” and “People who bought this also bought this” etc.

Mckinsey analysts estimated that 35% of AMAZON revenue and 75% of what they watch on NETFLIX came from product recommendations.[1] With the development in the recommended techniques and approaches, more RS has been implemented and many real-world recommendation applications have been developed. RS application includes recommending movies, music, TV program, books, websites, conferences, and learning materials.[2]

Objectives

- True reflection of the user experience.
- Enough spread for allowing decision making.
- Implementation of the RS increases revenue by 10-30%
- Recommendations can boost profit margins by recommending either high margin items or loss leader products or overstocked items.

II. ISSUES AND CHALLENGES

1) *Cold Start Problem* - it refers to a situation where the recommender engine will not be able to make any predictions of the users or items due to lack of initial tool usage or rating. This situation usually occurs when a new user or a new item is added to the system. Recommendations are not made as soon as the user rates an item. As the number of ratings provided by the user is insufficient to make a good prediction. Hence the probability of a new user leaving the system is high. This problem can be solved by using hybrid approaches.

2) *Scalability* - with the increase in the number of users, items and ratings, the system needs more information for processing information and making predictions. The majority of the resources are consumed for the purpose of determining the users with similar behavior and goods with similar descriptions. One of the solutions is to use an online learning algorithm which processes, each user sequentially and immediately and another solution is to use a distributed algorithm where the computations are done in parallel with multiple machines.

3) *Sparsity* - is encountered when there are large ,number of items to be rated, but the users would rate only a few of them. So when a user- item matrix is generated, only few entries would be present, which causes the matrix to be sparse leading to poor recommendations. Sparsity is a problem of lack of information. Demographic filtering can be used to overcome this problem.

4) *Overspecialization* - the items which are rated the highest are given as recommendations to the users. They might lose interest in the system as they have already bought and experienced the item. There is a high probability that the user might leave the system as it is of no use to the user. The problem can be solved by using neighborhood based collaborative filtering, genetic algorithms or by eliminating similar items.

5) *Privacy* - In order to provide accurate and correct recommendations, the system must obtain most of the information about the user and their location, including the demographic data. Questions related to the reliability, security and confidentiality arises. Many online stores offer effective protection of the privacy of users by utilizing specialized algorithms and programs.

6) *Trust* - the voices of the users with shorter histories may not be appropriate when compared to the voices with a rich history in their profiles. Trust issues arise during evaluation of customers. Distributing priorities to the users solves the issue.

III. BENEFITS OF RECOMMENDATION SYSTEM

There are two major benefits of using a recommender engine – revenue and customer satisfaction.

1) *Revenue and sales increase* - For years, the increase in the revenue has been the most popular indicator for every business owner. The successful recommendation systems implemented in the Amazon lead to the 29% increase in the annual sales. Amazon analyzed the most common purchases made by the users in order to obtain the insights their business intelligence systems did not find. It is impossible and complicated to analyze the list of purchases made manually, as it has a vast amount of data, it takes a long time to find correlations using traditional data analysis algorithms. Amazon uses machine learning algorithms to process the data. This resulted in the prediction of user purchases, according to their shopping history and recommend the next product.

2) *Customer satisfaction* - most of the users expect their recommendations to be predicted based on their browsing history as they think they would find better opportunities for good products. This could further help and guide in their online shopping or purchasing activities.

3) *Personalization* - People usually take suggestions from their family or friends, thinking they would know their choice better. Hence, they are good at recommending products and this is what RS try to model. One can use the data accumulated previously such as the user's previous browsing history so that the website's overall services can be improved and ensured that the recommendations are made as per his preferences. In return the user will be placed in a better mood to purchase the products and services.

4) *Discovery* - Flipkart, Amazon, Facebook and many other networking sites, make surprising recommendations which are similar to what one already likes. People usually expect they are recommended with items they like and when they use a site which can relate to their choices perfectly , then the user is bound to visit the site again.

5) *Provides Reports* - Is an integral part of a personalization system. Giving the customer accurate report allows him to make solid decisions about the site and the direction of a campaign. Based on these reports customers can generate offers for slow moving products in order to create a drive in sales.

IV. HOW DOES A RECOMMENDATION ENGINE WORK?

The four phases involved in the recommendation engine for processing data are as follows

1) Data collection - Data gathering is the first step in creating a recommendation engine.

Data is of two types – implicit and explicit data

a) Explicit data refer to the feedback data provided by the user in the form of ratings and reviews on products.

b) Implicit data refer to the data collected based on the user behavior like the browsing history, cart events, page views and search logs. This data will be created for every user visiting the site.

The data on the user behavior is easy to collect as he can maintain a log of his activities on a particular site. This does not require any extra action from the user. The downside of using the implicit data approach is that it is harder to analyze the data.

As each user has different insights about a product, their data sets will be distinct. Over time, as the engine is fed with more data, it gets smarter and smarter with the recommendations.

2) Data storage - More the data available to the algorithms better the recommendations will be. The type of storage needed can be decided based on the type of data used to create recommendations. For the purpose of storage one can use NoSQL database, a standard SQL database or of some kind of object storage. Each of these options is viable depending on whether the data captured is an implicit or an explicit data and also depends on factors such as portability, ease of implementation, the amount of data the storage can manage and the integration with the rest of the environment. When saving the user feedback, a managed database minimizes the number of tasks required and helps to focus on recommendations.

3) Data analysis - The items that have similar user engagement data is analyzed by filtering out the data using different analysis methods. Some of the ways in which the data are analyzed for immediate recommendations are shown below:

a) Real – time systems - The system involves tools that can process and analyze streams of information. It processes the data as it is created and provides immediate recommendations.

b) Batch Analysis - Requires the data to be processed periodically. In order to make the analysis relevant enough data needs to be created like the daily sales volume.

c) Near-real-time analysis - Allows the users to gather the data quickly so that they can refresh the analytics often (every few minutes or seconds). This system works best for providing recommendations during the same browsing session.

4) Data filtering - The next step is to filter out the data so that the necessary relevant data is obtained to provide the recommendations to the user. An Algorithm has to be chosen that would better suit the recommendation engine from the list of algorithms available.

V. TYPES OF RECOMMENDATION SYSTEM

The RS is classified into two categories:

1) Traditional Approach
2) Modern / Non – Traditional Approaches

Traditional approaches are further classified as:

1) Collaborative Filtering
2) Content Based Filtering
3) Hybrid Approaches
4) Knowledge based
5) Demographic approach

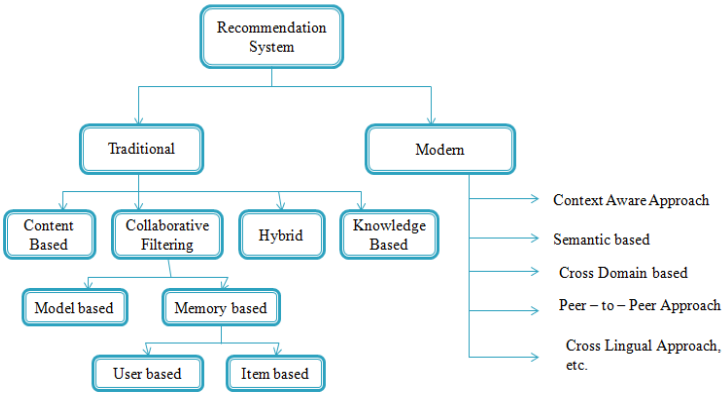

Figure 1: Classification of Recommendation System

Content based and collaborative filtering is explained in detail with a case study. Consider the scenario of a website that sells smartphones. With the growing amount of customers, the task is to showcase the best choice of smart phones to the customers based on their taste and preference. To understand how the engine works, let's consider a set of 5 smartphones having a battery and display as the major features.

- S1 - good battery life, but poor display
- S2 - amazing battery performance, but very rough display
- S3 - battery is one of the best, but display lacks quality
- S4 & S5 - good in terms of display, but poor in terms of battery performance.

Using the above characteristics an Item-Feature Matrix can be created. The value in the cell represents the rating of the smartphone based on their feature out of 1.

Table 1: Item-Feature Matrix

Smartphone	Battery	Display
S1	0.9	0.1
S2	1	0
S3	0.99	0.01
S4	0	1
S5	0.1	0.9

The sample set also consists of four active users with their preferences.

- **Amman**: He prefers battery over display as an ideal smartphone feature.
- **Bob**: He likes a long lasting battery.
-
- **Chandan**: He likes decent display with normal battery.
- **David**: He prefers display to be extraordinary, but not the battery.

Using the above data, a User-Feature Matrix is created as shown below:

Table 2: User-Feature Matrix

User	Battery	Display
Amman	0.9	0.1
Bob	0.8	0.2
Chandan	0.1	0.9
David	0.01	0.99

Having the Item-Feature and User-Feature matrices, recommendation of smartphones can be created for the users using the following algorithms.

5.1 Content based Recommendations.

Content based systems recommend items based on a comparison made of the similarity between the content of the items and the user profile. A user-item similarity is obtained by mapping the feature of items to the feature of the user. The highly matched pair is given as recommendations, as demonstrated below:

Here every user is represented by a feature matrix

Table 3: Feature Vector for users

	User	Feature Vector
U1	Aman	[0.9 , 0.1]
U2	Bob	[0.8 , 0.2]
U3	Chandan	[0.1 , 0.9]
U4	David	[0.01 , 0.99]

Every Item is represented as a Feature vector

Table 4: Feature Vector for items

Smartphone	Feature Vector
S1	[0.9 , 0.1]
S2	[1 , 0]
S3	[0.99 , 0.01]
S4	[0 , 1]
S5	[0.1 , 0.9]

Content Based ***Item User Mapping*** are given by the equation:

$$\text{MAX} (U_{(j)}^{T} . I_{(i)})$$

i,j -> n,m

For User U_1 (Aman), Smartphone recommendation is:

$\text{MAX}(U_1^T S_1, U_1^T S_2, U_1^T S_3, U_1^T S_4, U_1^T S_5)$

MAX([0.9 0.1]T [0.9 0.1], [0.9 0.1]T [1 0], [0.9 0.1]T [0.99 0.01], [0.9 0.1]T

MAX(0.82 , 0.9 , 0.89 , 0.18 , 0.10)

= S2(0.9), S3(0.89) & S1(0.82)

S2, S3 and S1 are recommended to Amman as they have the highest recommendation scores.

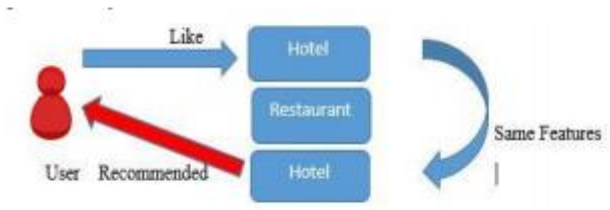

Figure 2: Content based Filtering

Merits

1. Other user's data not required.
2. No data sparsity as well as cold start.

Demerits

1. Essential to define the item features.
2. The excellence of the product can't be estimated..

5.2 Collaborative Filtering

User Behaviours are considered for recommending items in the Collaborative filtering algorithm. In order to recommend items to the new users, other users behaviour and preferences are used. Here, the item's features are not known.

Table 5: User-Feature Matrix

User	Battery	Display	Feature Vector
Aman	0.9	0.1	[0.9 , 0.1]
Bob	0.8	0.2	[0.8 , 0.2]
Chandan	0.1	0.9	[0.1 , 0.9]
David	0.3	0.7	[0.01 , 0.99]

Table 6: User-Behaviour Matrix

Smartphone	Aman	Bob	Chandan	David
S1	5	4.5	?	?
S2	5	?	0.5	?
S3	?	4	0.5	?
S4	?	?	5	4
S5	?	?	5	4.5

Where the values of the behaviour matrix can be described as:

Bi,j = {r , if rating is given ?, if no rating is given }

The user behaviour matrix can be used to derive unknown features of the most rated items
S1 is rated 5 by U1
S1 is rated 4.5 by U2
S1 rating by U3 & U4 are not known

Using the information feature vector:

S1 : [x1 x2] and the equations are:
U1TS1 = 5
U2TS1 = 4.5
[0.9 0.1]T [x1 x2] = 5
[0.8 0.2]T [x1 x2] = 4.
[0.9 * x1] + [0.1 * x2] = 5
[0.8 * x1] + [0.2 * x2] = 4.5
solving these equations, gives x1 = 5.5 and x2 = 0.5
S1 = [5.5 , 0.5]

Similarly,

S2 = [5.5 0]
S3 = [5 0]
S4 = [0.5 5.5]
S5 = [2.7 5.25]

Since, all the feature vectors are known, recommendations will be the mappings of user and item feature vectors. Thus for Amman, the recommendations as per is behaviour and preferences will be as shown:

MAX(U1TS1, U1TS2, U1TS3, U1TS4, U1TS5)
0.1
MAX([0.9 0.1]T [5.5 0.5] , [0.9 0.1]T [5.5 0], [0.9 0.1]T [5 0], [0.9 0.1]T
[0.5 5.5],[0.9 0.1]T [2.7 5.25])
MAX(5, 4.99, 4.95, 1, 2.9)

> S1, S2 and S3 - Since S1 and S2 have already been rated by him, S3 will be the new product recommended to Aman.

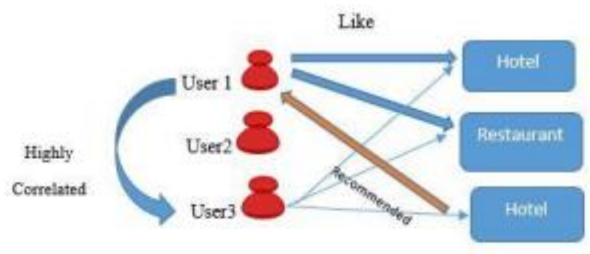

Figure 3: User based Collaborative filtering

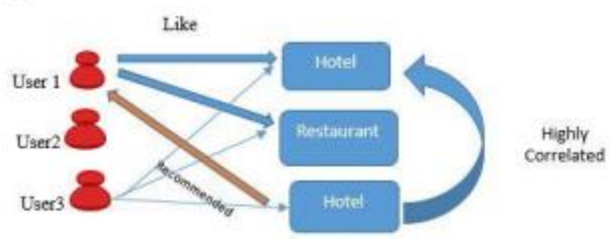

Figure 4: Item based Collaborative Filtering

Merits

1. Item's excellence is estimated through user ratings.

Demerits

1 Cold start problem
2 Stability vs. plasticity issue

5.3 Demographic filtering method

The Demographic filtering method uses users personal data such as age, gender, employment status, location etc. for recommendations. Its main aim is to classify the user based on the personal attributes and make recommendations based on demographic classes.

Merits

1 It is domain independent as the Item feature is not needed.

Demerits

1 Collection of demographic information, gives rise toprivacy issue.
2 Plasticity vs. Stability issue.

5.4 Knowledge based Filtering

Knowledge based system offers recommendations that already exist in databases or knowledge bases, that are not dynamically influenced by the ratings or recent preferences. The system is divided into two subcategories:
• Case-Based Systems - CBR are based on case-based reasoning, which relies on the similarity between a current case and the solutions that already exist in a database. The interaction with the system consists of four steps: Retrieve, Reuse, Revise and Retention.
• Constrained-Based Systems – Depend on the set of preferences, provides a set of possible solutions including explanations as to why these solutions were selected. A Constrained Satisfaction Problem is defined in Constraints set can be of three types: Compatibility Constraints, Filter conditions and Product constraints.

5.5 Hybrid System

According to the recent research, combining Collaborative and Content based filtering provides more effective recommendations. Hybrid approach recommendations can be made by implementing Collaborative and Content based predictions separately and then combining them. Several studies have compared the performance of hybrid systems with Collaborative and Content based methods and demonstrated that hybrid approaches provide more accurate recommendations. It can be used to overcome the common problems in RS such as cold start and the data paucity problem.The various other types of hybrid algorithms are weighted, switching, cascade, mixed hybridization etc.

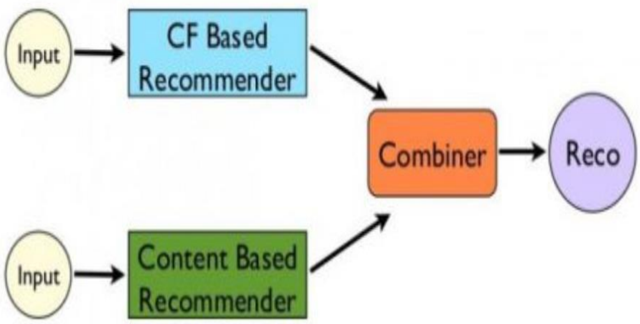

Figure 5: Hybrid Recommendation System

VI APPLICATIONS

- E-government recommendation system (Government-to-citizen and Government-to-business service recommendation)
- E-business RS
- E-Commerce or E-Shopping RS
- E-Library RS
- E-learning RS
- E-tourism RS
- E-resource service.

VII CONCLUSION AND FUTURE WORK

Recommendation systems open up new options for obtaining personalized information on the internet. It also helps to solve the problems of information overload which is a common aspect with the information retrieval systems and allows the users to have access to products and services which are not readily available to the users on the system. This paper discussed the various challenges, benefits, working, applications and different types of traditional recommendation algorithms with case studies, merits and demerits. This paper provides the basic information one has to know for developing a recommendation engine.

There are more advanced and non-traditional methods to power the recommendation process. Some of the techniques are deep learning, social learning and tensor factorization are based on the machine learning and neural networks. Such cognitive computing methods take the quality of recommendations to the next level. The product recommendation engines to improve the use of machine learning and create a much better process for customer satisfaction and retention.

REFERENCES

[1] Singh, G,Boparai, R S. A Survey on Recommendation System. ISOR Journal of Computer Engineering (ISOR-JCE) e-ISSN:2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 6,Ver. V(Nov-Dec. 2015), PP 46-51

[2] Isinkaye, Folasade & Folajimi, Yetunde & Ojokoh, Bolanle. (2015). Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal. 16. 10.1016/j.eij.2015.06.005.

[3] P V, A. and Joseph, D. (2017). A Survey on Recommender System Types And Its Classification. International Journal of Advanced Research in Computer Science. Volume 8, No. 9,November-December .