A COMPARATIVE STUDY ON RECOMMENDATION SYSTEM USING HYBRID APPROACH

Dr. P.V.R.D. Prasada Rao  
Professor, Department of CSE, KLEF, A.P, India

Vineetha B  
Student, Department of CSE, KLEF, A.P, India

Ch. Priyanka  
Student, Department of CSE, KLEF, A.P, India

V. Satish  
Student, Department of CSE, KLEF, A.P, India

ABSTRACT

Recommendation is a wide used technique to guide user to settle on the proper product on-line because it becomes the essential feature for a better E-commerce. Many recommender applications use item to item collaborative approach, i.e., overlapping previous user’s data with new user’s data and recommending remaining items. An item can be recommended to the user based on user’s interests, favorite subjects, occupation of the customer, person’s gender, age, etc. This literature survey tries to explain about item to item collaborative approach along with new features which improves the performance of the system. Not only in E-commerce, in many web applications recommendations play an important role such as book recommendation for a library, research paper recommendation based on previous searches while searching for a paper, dynamic recommendation system using web data mining and on social websites.

Key words: Recommendation system, Collaborative filtering, data mining techniques.


1. INTRODUCTION

Now-a-days many people are connected to internet, so the usages of internet for daily transactions are growing rapidly. Many E-commerce websites are emerging which in turn
changed the means of doing business in a traditional manner. This overnight extension has resulted in, new challenges to both corporations and to customers. Some people spend more amount of time for searching a product in online websites, and sometimes customers will have more choices which results in a lost state and confusion. It has become a major problem to find what product the user is searching for. A well solution to overcome this problem is recommendation system that provides and advises the customer with the type of items they are interested in i.e., suggesting a book to a student, for an example a student's favorite subject is chemistry and purchased some products in that journal, now this recommendation system suggests a chemistry book which is not purchased by him.

Many social networking websites has become a part of our life. Due to this, large amounts of transactional data and interactional data is produced day by day. So, to get some productive information from this data, recommendation systems are introduced. Data gathering and storing technology has made it probable for organizations to gather large amounts of data at less cost. Exploring this warehoused data, in order to extract suitable and accurate information, is the main goal of the general activity named as data mining. Data mining is a technique which is used to extract useful information from huge amounts of data. Recommendation systems use many data mining techniques for predicting a product to a customer in which they are interested in. Algorithm's such as classification techniques, frequent pattern mining, clustering techniques can be used.

Recommendation are generally classified into content and collaborative based approaches as follows-

![Figure 1](image)

**Figure 1** Different techniques of recommendation systems

Many famous websites such as amazon, ebay, netflix, youtube, facebook, twitter use recommendation systems. Not only in E-commerce, these recommendation systems can be used in E-learning systems. Recommendation System in E-Learning is that system that suggests the scholar, the course to be opted based on the previously gathered data about their choice for selected set of subjects e.g. a student who is presently studying the course “Database Management System” may like to study the course which is similar to the previous one i.e., “Advanced Database Management System”. These Recommendation systems can also be used to recommend a book, for example: A person purchased a book named” Half Girl Friend” which is written by Chetan Bhagat, now using recommendation systems recommend another book which is written by Chetan Bhagat to that user because many users buy books in the journal they are interested in. Recommendation systems can also be used in applications like bookmyshow to recommend a movie to a user based on their previously watched movies for example, let us consider a person who always watches horror movies such as Conjuring, horror etc, this recommendation systems searches for the movies of similar type by comparing with the movies which are not yet watched by the user such as “It” and recommends that to
the user. In this way, Recommender systems are having many and many applications. In the beginning stage, recommendation systems recommend a product or a service to the user based on the ratings given by that user for different products. Recommendation systems may be considered as a greater extension of conventional information systems used in commerce such as travel, hospitality and in many businesses. However, these recommendation systems have precise roots and are more similar to artificial intelligence (AI) than other Information Technologies. A recommendation system studies from a purchaser’s behaviour and suggests a product in which the users is interested. Recommender systems recognize recommendations separately for individual users based on earlier purchases and explorations. In general, Recommendation systems are divided into two types-Content based and collaborative based systems [1]. Content based systems mainly concentrates on the properties of the product to make recommendation, Collaborative filtering concentrates on user and item interaction to produce recommendations.

2. LITERATURE SURVEY

Yingtong Dou [1] mainly focused on analysing data of social networking sites that generate a huge amount of data daily. The concepts used here are user based collaborative filtering, item based collaborative filtering. User based collaborative filtering takes the data of different users who are similar based on the ratings given to the products and predicts the rating for an unpurchased item and recommends it to the user. To find similarity between the users, k nearest neighbors algorithm is used. In item-based filtering, rather than comparing users, it compares the similarity of many items and then suggests that item by predicting the rating to that item. This paper also explains about different challenges faced while building a recommender system. Here drawbacks are cold start problem and sparsity problem and this method might be costlier sometimes.

In the paper written by Prajyoti Lopes [2], a system that dynamically makes recommendation as per altering behavior of users with the use of Web Usage Mining. In Web Usage Mining, user’s click stream data plays a huge role, this Clickstream data consists of user’s path over a website. This data is stored in web log files which is highly unstructured data. This data is pre-processed, and recommendation techniques are applied. In general, there are two types of errors in recommender systems, false negatives and false positives. False negatives mean user is interested in that product, but recommender system fails to recommend. False positives mean user is not interested in that product, but recommender system recommends that product to the user. The proposed system in this paper reduces the false positives that occur regularly in conventional recommender systems.

In the paper written by Paritosh Nagarnaik[3] Web mining plays a major role for discovering the common data pattern from Internet websites. In this paper, web page recommendation is done using hybrid collaborative filtering approach to provide recommendations even for novice users. It is a combination of both content and collaborative approach. Data is collected from various sources and it is pre-processed. Now clustering algorithm such as k-means is applied to make clusters of similar users. Now on each cluster CHARM association rule algorithm is used to find the frequent patterns, Finally, based on order of earlier searches next page is recommended. This hybrid based approach advances the quality and accuracy of recommender systems.

Sunita B Aher[4] concentrates on recommending a course to the user or student who is using E-learning system, e.g. a student who is studying the course of “Database Management System” may like to study the course which is similar to the previous one i.e., “Advanced Database Management System” The algorithm proposed in this paper is a combination of
Simple k-means clustering and Apriori association rule algorithm is used. This system is also applicable to online shopping. K-means is an unsupervised technique which clusters similar type of subjects into a set. Apriori is used to mine the repeated pattern in the database. In the combined approach first clustering is applied and each cluster is stored in different files, then for every individual cluster apriori is applied to find frequent patterns. Rather than using only apriori algorithm, this combined approach gives improved results to recommend a course to the user of E-learning Systems.

Ayhan.demiriz[5] explains about the experimental study on both internal data and publicly available data using sparse binary data. They have proposed a new recommender system, E-Vzro which uses various association mining techniques. The main advantage of this paper is that the method of E-Vzpro can be used to perform sequence mining and one of the most important benefits of this recommender system is, it can be used to predict cancellation and offer retention to the customers.

The E-Vzro ‘s training time is very insignificant. This implementation has a drawback .i.e, it can’t predict longer prediction times. For example, the conduct time between the order of a car and delivery of a new car from a manufacturer may be from one month to twelve months. While future work of this paper, while at the beginning of the recommender system there might be problematic due to the unavailability of user history in such cases recommendation varies at each time when the customer tries to interact with the system. In such cases, the recommender systems should also avoid null predictions.

In this paper Gediminas Adomavicius1 and Alexander Tuzhilin [6] describes about the current generation of recommender systems techniques. This are classified into three main categories: content based, collaborative based and finally hybrid based recommender systems approaches .This paper describes about various limitations and describe various extensions that can make recommender systems more applicable in various fields in a better way with more accurate result. This extension includes various information regarding the customers and their items.

In this paper, they have researched about New User Problem-where the customer is new to the online marketing, the system must provide correct recommendation by using previous ratings, clicks, likes, dislikes and feedback.

New Item Problem-In this problem the item is new to the market, so we must be careful while recommending to the customer until and unless it is liked or clicked by the customer.

In the paper V.Kumutha and S.Palaniammal[7] describes more about the high dimensional data in various areas. As the dimensions increases the number of clustering techniques are increased for the quality of results. While in high dimensional data, the data is scattered, and the distance measures becomes useless. This problem has been studied carefully and in an effective way that covers a large area and found various solutions.

Subspaces clustering integrates evaluation and clusters in order to find different clusters in various subspaces. While they have used various algorithms like Subclu, Proclus, fires, Inscy and Doc clustering techniques are used and compared based on accuracy and coverage.
Table 1 Comparison of Different algorithms

<table>
<thead>
<tr>
<th>ALGORITHM</th>
<th>ADVANTAGES</th>
<th>DISADVANTAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>easy to recognize and implement.</td>
<td>If data is changing frequently, then results are not precise.</td>
</tr>
<tr>
<td>Association Rule</td>
<td>Uses current data and predicts which gives exact results and easy to implement.</td>
<td>Number of lookups into the database are more.</td>
</tr>
<tr>
<td>Or item to-item collaborative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision tree</td>
<td>Predicts exactly even if the data is non-linear.</td>
<td>For categorical data, number of levels of the tree will be more which makes more complex to calculate the results.</td>
</tr>
<tr>
<td>Fp-Growth</td>
<td>Number of database scans are less.</td>
<td>If data is not possible to divide, then this method fails.</td>
</tr>
<tr>
<td>K-Means</td>
<td>Works good for larger data also.</td>
<td>If data contains outliers, prediction goes entirely wrong.</td>
</tr>
<tr>
<td>Apriori</td>
<td>Easy to find frequent item subsets.</td>
<td>More Search space and cost</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of lookups into the database are more.</td>
</tr>
</tbody>
</table>

In the paper written by Basiliyos Tilahun Betru[8], different methods of recommendation system techniques such as content based, collaborative filtering and hybrid based methods are explained. Collaborative filtering is further classified into model based and memory based methods. A new approach named Deep learning approach is introduced which is new area for machine learning. Deep neural networks which consists of several layers that are hidden and fully connected is used. Due to the flaws of the conventional recommendation methods in gaining accurate result a deep learning method is presented which enables the model to learn about different kinds of users and items inevitably to improve accurateness of recommendation. With the help of advantage of deep learning in learning many types of data, deep recommender systems will understand customers demand and advance the quality of recommendation. This paper is a initial study of the deep learning approach for recommendation systems which can be improved further.

Daniar Asanov[9] mainly concentrates on movie recommendation systems and different methods that are used to build a movie recommender system. Apart from traditional approaches like content based, collaborative and hybrid based approaches new modern approaches like context aware approach which recommends product based on the environment in which the user lies, semantic based approach which classifies the users and items regarding to their domains and then recommends by applying data mining techniques, Cross domain based approach which specifies that a user may have similarities with other in one domain but in other domains they may or may not have similarities. So, in cross domain systems recommendations are dependent on domains. Peer-to-peer approach, in this approach each person can compare itself to a group of other persons with similar taste, interests and can get recommendations from that group users. This paper also explains about the challenges such as cold start problem, sparsity problem, trust, scalability, privacy problems faced during the development of recommendation systems.

In the paper written by Greg Linden[10] focused on about most target marketers; the recommender algorithms provide a very efficient form of marketing by creating a personalized shopping experience for every customer. This is all about large retailers like flipkart, Amazon uses a better recommender algorithm that is scalable over a huge customer bases and product catalogues that requires only sub second processing time that is taken to generate recommendation through online. This recommendation is able to react to slight change in user’s data and recommend without regarding the number of purchases and rating.
This kind of recommendation problem is overcome by using item-to-item collaborative filtering technique, by using this technique we can expect our retail industry to mainly apply recommender system algorithms to reach target marketing secrets in both online and offline.

J. Ben Schafer [11] describes about popular Collaborative filtering based algorithms which produces meaningful results in various domain’s. The data mining techniques are used in both hybrid systems, that improves the recommendation in an application where the customer can get correct results of his taste. This use of data mining techniques has changed the types of recommendations i.e., when to recommend and where to recommend. These types of techniques have increased the marketing strategy.

PVRD Prasad Rao, K Divya Bharati [13] provided a new effective method that works well in heterogeneous architectures. Now-a-days, many software’s are in heterogeneous environment, if recommendation systems are used in such environment we can use SAVE framework to implement it with dynamic resource utilization.

3. OVERALL SURVEY:

This survey helps to propose a model that helps to recommend a product in E-commerce websites, to recommend a course in E-learning systems and in movie recommendations, music applications etc. To implement this model first we need to collect transactional and interactional data from various resources and pre-processing techniques must be applied so that the data is in a structured way without any duplications and errors.

![Combination of k-means and fp-growth](image)

**Figure 2** Combination of k-means and fp-growth.

On this data Clustering techniques such as k-means, k-medoids are applied to find the similarities. Now, on each cluster to find the frequently repeated patterns a priori algorithm or fp-growth algorithm are applied. New researches suggest deep neural networks which is similar to machine learning algorithms. Even though it has an excessive impact in many areas, more improvement need to be done in applying this model to recommender systems.

4. CONCLUSION & FUTURE SCOPE

Now-a-days recommendation systems play a key role in marketing strategy because of increasing online sales. This changed the way of doing business in a conventional way. In this paper, different recommendation techniques are described. Many clustering, classification, association algorithms are used in building recommendation systems, rather than using individual algorithms if a combination of algorithms (hybrid approach) are used,
recommendations can be predicted better. So, from this research we can conclude that combination of algorithms such as k-means and apriori can give an improved result.

Future scope involves in discovering an advanced combination of data mining techniques that provide good results than existing combination. Even deep learning methods can be improved which delivers better recommendations.

REFERENCES


