A Survey on Recommender Systems

Aakash Sarnobat
Computer Engineering Department
Shah & Anchor Kutchhi Engineering College
aakashsarnobat@gmail.com

Dhara Kalola
Computer Engineering Department
Shah & Anchor Kutchhi Engineering College
dharapatel23@gmail.com

DOI: 10.29322/IJSRP.9.09.2019.p9356
http://dx.doi.org/10.29322/IJSRP.9.09.2019.p9356

***Abstract*—Recommender systems provide user with facilities that the user might be interested in. These systems provide options based on the pattern of usage, certain information and reference with certain characteristics to present recommendations to the user. This paper describes the classification of different recommender systems along with the limitations of each. It also covers some real world applications of recommender systems and possible systems that can be built using recommender system algorithms.**

***Keywords*—Recommender System, Content Based Filtering, Collaborative Filtering**

INTRODUCTION

Recommender systems are online information systems that recommend products to customers by performing some sort of automatic selling in e-commerce, or more generally, recommend information items to users based on their taste. Although people's tastes vary, they do follow patterns. People tend to like things that are similar to other things they like as well as another similar behavioral person likes. e.g., books to library users, research papers to cite seer users, courses to students etc. Recommender systems typically are personalized and targeted at the individual customer. To achieve personalization, they make use of customer profiles consisting of explicit and implicit product ratings. In e-commerce businesses, recommender systems enhance revenues, for the fact that they are effective means of selling more products. In computer science, the term recommender system often is used synonymously with collaborative filtering, in the context of information retrieval they are also known as relevance feedback. [2]Recommender systems come under the category of information filtering system that attempt to predict the 'rating' or review' that a user would give an item.

CLASSIFICATION OF RECOMMENDER SYSTEMS

There are two broad categories of recommender engine algorithms: user-based and item-based recommenders. These are based on two filtering techniques which are collaborative and content-based filtering. Collaborative filtering produces recommendations based on, knowledge of user's relationship with items, products and services. However, this technique reveals two major issues: sparsity problem and the scalability problem. On the contrary, CB analyses documents rated by an individual user and uses the contents, as well as the provided ratings, to infer a user profile that can be used to recommend additional items of interest.

A. Content Based Filtering

Content-based filtering recommender systems are those which recommend items based on contents of items rather than users rating of the system. Instead of using a user-to-item correlation and defining methodologies, they make use of item-to-item correlation for generating the recommendations. The steps carried out in the process of generating recommendations are as follows:

- Gather information about the item (For example- title of the book, author, cost etc. are some of the contents.
- Process data and churn out useful features and about its content.

Content-based technique is a domain-dependent algorithmic technique and it emphasizes a lot on the analysis of the attributes so as to generate predictions. When documents such as web pages and publications and are to be recommended, content-based filtering is the most successful. In content-based technique, recommendations are made based on the profiles using features that are extracted from the contents of items which the user had earlier evaluated.

In a content-based recommender system, the items and a user profile is built using the keywords described. It indicates the type of item the user likes. These algorithms recommend

items that are similar to those which a user liked in the past. Various candidate items are compared with the items which were previously rated by the user and the best-matching items are recommended. Content based filtering has its roots in retrieval of information and information filtering.

B. Collaborative Filtering

Another approach to the design of recommender systems that has wide use is collaborative filtering. This approach includes recommendations which are based on a few customers who are most similar to other active users. Collaborative methods are based on collecting a large amount of information and analyzing it based on users' behaviors, activities or preferences. It then predicts what users will like based on their similarity to other users. A key advantage of the collaborative filtering approach is that it is independent of machine analyzable content and so it accurately recommends complex items like without requiring an understanding of the item itself.

Different algorithms have been used in measuring user similarity or item similarity in recommender systems. The collaborative filtering can be revamped with neighborhood methods, whose focus is on the relationship between the items or, between the users. They are:

1. *User-based CF*: For each user, it computes the correlation with other users. For each item, the rating of the users which are highly correlated with each user are aggregated.
   Problem: Sparsity, easy to attack
2. II. *Item-based CF*: For each item, it computes the correlation with other items. For each user, his rating of the items which are highly correlated with each item is aggregated.

C. Hybrid Recommender Systems

Hybrid recommender systems is another category of recommender system which strives to overcome the limitations of the other approaches discussed before. It is a combination multiple recommendation techniques. The most widely used hybrid approaches are those of content based and collaborative filtering. It uses both item content and ratings of all users.The idea behind hybrid techniques are that a combination of algorithms will give more accurate and effective recommendations than a single algorithm. The disadvantages of one algorithm can be overcome using another algorithm. Multiple recommendation techniques can overcome the weaknesses of an individual technique. The combination of approaches can be done in the following ways:[5]

1. Implement collaborative and content-based techniques separately and combine their predictions,
2. Incorporate some content-based properties into a collaborative approach,
3. Incorporate some collaborative properties into a content-based approach, and
4. Construct a general unifying model that includes both content-based and collaborative characteristics.

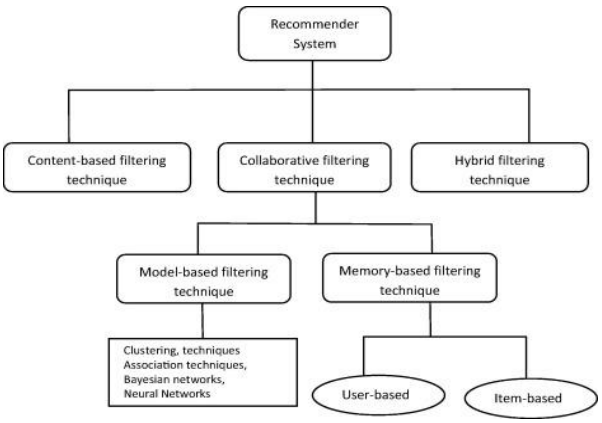

Fig. 1. Classification of recommender systems

II. Recommender System Examples

In the following section we present four e-commerce businesses that utilize one or more variations of recommender system technology in their web sites.

Amazon.com

We focus on recommender systems in the book section of Amazon.com.

Customers who bought: Like many E-commerce sites, Amazon.com™ (www.amazon.com) is structured with an information page for each book, giving details of the text and purchase information. The Customers who Bought feature is present on the information page for every book in their catalog. It is in fact two separate recommendation lists. The first recommends books frequently purchased by customers who purchased the book. The second list recommends authors whose books are often purchased by customers who purchased works by the author of the selected book. [4]

Eyes: The Eyes feature allows customers to be notified via email of new items added to the Amazon.com catalog. Customers insert requests based on author, title, subject, ISBN, or publication date information. For notification queries customers can use both simple and complex Boolean-based criteria. Requests can also be directly entered from any search results screen, creating a persistent request based on the search.

Amazon.com Delivers: This is a variation on the Eyes feature. Customers have an option to select from a list of specific categories (Science fiction books, biographies, cooking). Periodically the editors at Amazon.com send email announcements to notify subscribers of their latest recommendations in the subscribed categories.

Book Matcher: The Book Matcher feature allows customers to give direct feedback about books they have read. Customers rate books they have read on a 5-point scale from "hated it" to "loved it." After rating a sample of books, customers can ask for recommendations for books they might like. At that point a half dozen non-rated texts are presented which correlate with the user's indicated tastes. Feedback for these recommendations are given by a "rate these books" feature where customers can indicate a rating for one or more of the recommended books. Customer Comments: The Customer Comments feature helps customers to receive text recommendations based on the opinions of other customers. Each book has an information page which has a list of 1-5 star ratings and written comments provided by customers who have read the book in question and submitted a review. Customers have the option of taking into account these recommendations into their purchase decision.

III. Limitations of Recommender System

In content-based recommendation, the utility *u (a, i)* of item *f* for user *a* is estimated based on the utilities *u (a, fi)* assigned by user *c* to similar items.

Content-based recommender systems have several constraints that are described as follows-

A. Limited Content Analysis

Content-based approaches are restrained by the features that are clearly associated with the objects that these systems recommend. Therefore, in order to have a acceptable set of features, the content must of the form that can be parsed automatically by a computer (e.g., text) or the features should be given to items manually. While information recovery techniques work well in obtaining features from text files, some other domains have an inbuilt problem with automatic feature extraction. For example, automatic feature extraction methods are tough to apply to multimedia data like audio streams, and video streams.

Another issue with limited content analysis is that, if two different objects are represented by the same set of features, then they cannot be separated.

B. Overspecialization

When the recommender can suggest the items that are rated very high against a user's account, the user is limited to being suggested the items that are similar to those already rated. For example, a person who has never tried Italian food would never receive a recommendation for even the greatest Italian hotel in town.

C. New User Problem

The user has to rate a significant number of items before a content-based recommender can really understand the user's preferences and give the user with reliable suggestions. Therefore, a new user, having seldom ratings, would not be able to get accurate recommendations.

IV. Limitations of Collaborative Methods

Unlike content-based recommendation methods, collaborative recommender systems predict the utility of items for a specific user based on the things previously rated by other users. For example, in a movie recommendation system, in order to recommend movies to the user, the collaborative recommender system will try to find the "peers" of the user, i.e., other users that like similar movies (rate the same movies similarly). However, collaborative systems have their own limitations as described below.

A. New User Problem

The collaborative system has the same problem like the content-based systems. In order for the system to make accurate recommendations, the system will first learn the user's preferences from the ratings that the user will give. Several methods have been proposed to address this issue. Many of them use the hybrid recommendation technique, which combines both content-based and collaborative technique.

B. New Item Problem

New items are added on a regular basis to recommender systems. These systems rely only on users' preferences for recommendations. Therefore, the recommender system would not be able to suggest it, until the new item that has been added is rated by a considerable number of users. This problem can also be resolved using hybrid recommendation techniques.

C. Sparsity

In any recommender system, amount of the ratings obtained is usually small compared to the expected number of people. Accurate prediction of ratings from few numbers of examples is important. Also, the success of the collaborative recommender depends on the supply of a certain users. For example, in the movie recommendation application, there may be many films that have been rated by few people and these movies would not be recommended or sometimes it would be recommended, even if those some users gave high ratings to those movies.

Related Work

There are some papers are available in literature on big data analysis for recommender systems using several platforms:

Jai Prakash Verma [6] projected a recommendation system for the massive quantity of data available on the web in the form of ratings, opinions, complaints, and feedback about any item or service using Hadoop Framework. They have implemented Mahout Interfaces for analyzing the data provided by review and rating site for movies.

Deuk Hee Park [7] explores trends in recommender systems research and provides researchers with insight and future direction on recommender systems. They reviewed 210 articles from 46 journals published between 2001-2010 and categorize them into eight application fields.

Gediminas Adomavicius [5] discusses limitations of current recommendation methods and discusses attainable extensions to enhance these Systems. These limitations consists of an improvement in the understanding of products, combining contextual information, support for varied criteria ratings, and more versatile and less complex types of recommendations.

Conclusion And Future Work

Thus, this paper intends to provide an overview of some approaches of recommender systems along with providing their applications and limitations.

A medicine recommender system can be built for medical trainees and doctors wherein they can search for medicines based on the ingredients and manufacturer of the medicine. The dataset will contain a list of all drugs available along with their manufacturer and ingredients. The dataset will also contain symptoms associated with each drug. When the user logs into the system he will be asked to select the symptoms of disease or ailment. The user can also enter the ingredients to filter out the desired drug name from the entire list. This system can be build using the hybrid approach which will avoid the limitations of content and collaborative methods.

References

[1] https://en.wikipedia.org/wiki/Recommender_system
[2] Compound Classification Models for Recommender Systems Lars Schmidt-Thieme Institute of Computer Science, University of Freiburg, Germany
[3] Dr. Sarika Jain, Anjali Grover, Praveen Singh Thakur, Sourabh Kumar Choudhary, Trends, Problems and Solutions of Recommender systems.
[4] Greg Linden, Brent Smith, and Jeremy York, Amazon.com Recommendations Item-to-Item Collaborative.
[5] Gediminas Adomavicius and Alexander Tuzhilin, Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions
[6] Jai Prakash Verma, Bankim Patel, Atul Patel, Big Data Analysis: Recommendation System with Hadoop Framework, 2015 IEEE International Conference on Computational Intelligence & Communication Technology
[7] Deuk Hee Park, Il Young Choi, Hyea Kyeong Kim, Jai Kyeong Kim, A literature review and classification of recommender systems research