Hybrid Recommendation Engine using Feature-Combination

Samiksha S. Lohé

Department of Computer Science & Engineering, Government College of Engineering, Amravati, India

Abstract—An abundant amount of information is created and distributed over electronic media. Recommendation systems use various filters to extract required information. In every aspect of our life, recommendation engine is used to make such a task easier. We proposed a hybrid filtering engine for books data by using the combination of filters. We compared the proposed approach with existing approaches using algorithms, and found that our approach outperformed existing methods in terms of accuracy and robustness. Hybrid system combines the two or more filters to gain better accuracy and performance. The proposed recommendation engine is able to make a large number of recommendations simultaneously, and fit the partial dataset to the appropriate model to have better performance. The engine is able to determine user’s personality traits by utilizing the book’s features. Three types of scenarios are covered here. If user is new then recommendations are made depending upon user interests, second is recommendations based on past purchase history and the last is recommendation by using different algorithms and cosine similarity.

Keywords—Hybrid filters, Accuracy

Introduction:
The large amount of data is present on internet about specific products and services. And these generate confusion among choices. To deal with such tasks, we require a best solution. By studying the current behaviour and patterns shown by the user, a good recommendation model is able to predict the items that the user may like in near future. Recommendation systems are able to make recommendation to users about the product and services that they are highly likely to buy or use. In this paper, content-based and collaborative-based filters are used to build a model for recommendation. The primary goal of content-based methods is to try to build a model, based on the available “features”, that explain the noted user-item interactions. The content-based approach utilizes the additional information about users and/or items. In this filtering method, item features to recommend other items similar to what the user likes and also based on their previous actions or explicit feedback. The Collaborative filtering method for recommender systems is a method that is based on the past interactions that have been recorded between users and items, in order to produce new recommendations. Collaborative Filtering tends to find what similar users would like and release the recommendations in order to classify the users into clusters of similar types and recommend each user according to the preference of its cluster. Such memory-based approach suggests an item to user A based on the interests of a similar user B. We proposed a system which uses a model built by hybrid filters approach. Hybrid approach is a combination of filters to recommend the products and services; which is way better in case of accuracy and robustness. This system helps to overcome the limitations of traditional filtering approaches. These methods can also be used to improve the performance of recommender systems in order to reduce some of the common problems such as cold start and the sparsity problem, as well as the knowledge engineering in knowledge-based approaches. Lastly, it is possible to employ a hybrid algorithm, which combines content predictions using TF-IDF and collaborative filtering with similarity indexing.

Literature survey:
Recommender systems are software tools used to generate and provide suggestions for items and other entities to the users by exploiting various strategies. Hybrid recommender systems combine two or more recommendation strategies in different ways[4]. We address the most relevant problems considered and present the associated data mining and recommendation techniques used to overcome them [2]. We also explore the hybridization classes each hybrid recommender belongs to, the application domains, the evaluation process and proposed future research directions. Based on our findings, most of the studies combine collaborative filtering with another technique often in a weighted way. Also, cold start and data sparsity are the two traditional and top problems being addressed [1]. Machine learning implementation utilizes the user based collaborative filtering to design the recommendation model. Using this collaborative filtering-based model efficient recommendations are possible[8]. In this paper[10], author describes a generative adversarial network-based image retrieval system. This recommendation model is highly useful for ecommerce platforms. Content-based recommendation system recommends books to the buyers based on the contents of the buyers’ past book buying history. Buying history gives the content of the overview of the books, in which buyer is generally interested from the large number of books [5]. Content recommendation system filter books based on the content of the book, in which buyer is interested. Although, content-based and collaborative filtering are the most popular in the various applications, both of them suffer from several limitations [6]. Hybrid recommendation approach, as a combination of two or more approaches, has been proposed to overcome the main limitations of traditional recommendation approaches and improve the quality of the recommendation offered. Hybrid recommender systems combine two or more recommendation strategies in
Different ways to benefit from their complementary advantages[7]. This recommendation algorithm is based on user using a user rating of historical information to generate user interest and target the nearest neighbour set, and then according to the nearest neighbour set generate recommendations for target users[9]. Also, the proposed approach will improve the recommendation accuracy when there is comparison with the Collaborative Filtering using cosine similarity approach and it will be faster. To enhance the proposed method the machine learning and deep learning method could also be used[3].

**Methodology:**

1. **Content based filtering**
   Content based filtering use the data exclusively about the data items. For this we need to have minimal understanding of the books preferred by users, so that we can then recommend books with similar tags/keywords to those specific users. In this paper, we used TF-IDF vectorizer approach for this problem. This approach works by identifying the keywords to obtain a vector representation of the data.

   **How Tf-idf works**
   Tf-idf counts the no. of times a particular word occurs in a document and then compares this count to the no. of times the word shows up in a bunch of other documents in a collection. It then generates a rank for each word. The following equations need to be applied to train the model.

   
   $$\text{tf}(t, d) = \frac{f_d(t)}{\max_{w \in d} f_d(w)}$$

   $$\text{idf}(t, D) = \ln \left( \frac{|D|}{|\{d \in D : t \in d\}|} \right)$$

   $$\text{tfidf}(t, d, D) = \text{tf}(t, d) \cdot \text{idf}(t, D)$$

   $$\text{tfidf}(t, d, D) = \frac{\text{idf}(t, D)}{|D|} + \text{tfidf}(t, d, D)$$

   $$f_d(t) := \text{frequency of term } t \text{ in document } d$$

   $$D := \text{corpus of documents}$$

   This trained model works with keyword extraction; where it automatically extracts the most used and most important keywords from a text. It uses machine learning artificial intelligence (AI) with natural language processing (NLP) to break down text so that it can be understood and analysed by machines.

2. **Collaborative filtering**
   This filtering method analyses the nature of each item. In this case, we are recommending books to a user by performing cosine similarity on the content of each book. Collaborative Filtering does not require any information about the items or the users themselves. It provides options of books based on user’s past behaviour. Among the various types of collaborative filtering techniques, this system uses model-based method.
3. **Hybrid filtering**

A hybrid recommendation system is a special type of recommender system which is a combination of the content-based and collaborative filtering method. Here, hybrid approach is using feature combinations of data items to provide recommendations. Hybrid recommendation approaches have been proposed to overcome the limitations of traditional recommendation approaches and improve the quality of the recommendation offered.

![Flow diagram of working of collaborative filtering](image)

**Fig. 1:** Flow diagram of working of collaborative filtering

![Flow diagram of working of hybrid model](image)

**Fig. 2:** Flow diagram of working of hybrid model
Implementation of proposed methodology:

We take the help of hybrid filtering approach as we do not get the satisfied results from the CF and collaborative approaches. We also take the help of Web scraping through which books data from various sites which can be useful in giving recommendations. Sometimes, it may happen that the old rating can become redundant or rather obsolete which can have a hostile effect on the recommendations. Hence filtering of this old rating becomes the necessary step. The proposed system is Hybrid Recommendation system; which is designed to control the cold start issue and reduces dependency of rating-based system. User can search any books by its title or author name. At first, user is been asked to fill certain information like their category preferences, authors, title of book, etc. for finding similar users. Based on this information, system will recommend the books to users which in turn help to overcome cold start problem. After entering information, user can see their liked category and liked author’s books in different titled catalogues. They will be shown books based on the history as well as the information they’ve provided. User will see random recommendations and predictions using different algorithms like SVD, KNN, RBM and Hybrid model recommends the books based on their recent ratings. Furthermore, the system will track history of users and that will reflect latest recommendations for book recommender system. The Tf-IDf vectorization is used in content-based filtering which is applicable for the keyword extraction. And these flaws will reduce when train the model with algorithms used in collaborative-based filtering. Collaborative-based filtering is a domain independent prediction technique. Hybrid recommendation approaches have been proposed to overcome the main limitations of traditional recommendation approaches and improve the quality of the recommendation offered.

Proposed system:

Algorithmic approach:

1. Start
2. Load the book database and user interface module.
3. Enter the required information into the fields.
4. Content-based filtering: Filter Book database using Tf-IDf vectorizer and present the list in ascending order.
5. Collaborative filtering: Perform item based collaborative filtering on this list of books and present it in ascending order on the basis of the ratings of books.
6. Hybrid filtering: Find out the intersection of the step 4 and step 5 and arrange them in ascending order of recommendations using feature-combination.
7. The result of step 6 is the actual recommendation of books; which will display on window.
8. If user wants recommendation on other books, refresh the window or go to step 2. Otherwise go to next step.
9. Stop
Benefits of proposed system:

a. High accuracy
Since the data is reduced to small size by removing errors and anomalies which is efficiently divided into clusters based on user’s preferences for the recommendations hence there is more chances that recommended books will be more accurate.

b. Less Processing times
Data which is used for recommendation is filtered and is updated on periodic basis where the user recent searched, viewed and rated books are taken into consideration only and only popular books of other users are considered while recommending. In all modules processing the data sent to next stage is filtered due to which modules processing will be less which finally results in fast processing.

c. Less work to be done by the user
Since the recommendation gives the user the books which he/she may preferred to read or buy after reading their favourite books hence user directly don’t need to search for the books manually.

Conclusion:
This paper describes various filtering techniques to build a book recommendation engine. Algorithm presented here for proposed system will be the effective one in order to show the best values of accuracy measures. Hybrid recommendation approach has been used to overcome the limitations of traditional filters. A well-built book recommendation engine can lead to an excellent customer experience; hybrid approach can benefit from the strengths of both item-based and customer-based methods.

References: