A Hybrid Movie Recommender system based on Content and Collaborative Filtering methods

P. Sabari Ramkumar, J. Aswin,
Student, Student,
CSE,
[1]SRM Institute of Science and Technology, Chennai, India

Abstract: Recommendation systems play a important role in suggesting movies to its users. The problem faced in the current recommender systems is the perspective of the user keeps changing and the system cannot work initially due to lack of data. The proposed work is to solve the cold start problem and recommends movies to its users through user perspective and the items similarities A hybrid method is proposed where content and collaborative based methods are combined where content based technique uses a similarity based approach and for user based collaborative filtering a model based approach is used, and for item based collaborative filtering a neighbor based approach is used. By combining these filtering methods overall efficiency of the system is improved.

Index Terms - Streaming media, Cold start problem, Content based filtering method, Collaborative based filtering method

I. INTRODUCTION

Recommendation systems have become more common in the recent decade, and are being used in a variety of domains. There are many movies to search and its very difficult to choose which one is more suitable for us. Movie Recommendation systems helps us to filter from a pool of movies to the one's more suitable for us. Hence recommendation system should be in a way that it should recommend more accurate results.

II. LITERATURE OF SUREVY

Bagher Rahimpour Cami [2] proposed capturing the temporal preferences user in a content based filtering method which provides a user centered structure that uses the content attributes of previous movie ratings. Which then put to a Dirichlet Process Mixture Model to refer from user preferences and give a proper recommendation of movies.

Yeo Chan Yoon [3] proposed a machine learning approach to recommend movies to users using K-means clustering algorithm to separate similar users and thus creating a neural network for each cluster.

Shreya Agrawal [8] proposed a more efficient method for improving scalability. This has made the computation time to be very less using cosine similarity.

Ningning Yi [4] suggested that continuous improvement in the field of Internet technology, data overload is becoming a huge issue. Getting some useful data from the system is becoming harder. Although a search engine can be of some use to the users find relavent data they need from the large amount of information to an particular limit, but it cannot solve the problem of data overload completely, when the users cannot describe the particular data they need, we need a recommender system to help the users to find the needed data.

Kunal Shah [6] gives an overall view of various recommendation systems and summarizes the present generation of recommendation systems and proposes recommendation systems is a subset of information filtering system. They attempt to put out various limitations of recommendation systems and their advantages. And suggests that hybrid recommender systems provide a notable improvement in accuracy, precision.

Jeffrey Lund [5] proposes a proposes an approach based on deep learning using artificial neural networks to produce a collaborative recommendation system which analyses the respective movie ratings for an end user, on the basis of a larger database of ratings. This system found that it performs better than a user based method on the basis of root mean squared error value evaluation on the ratings which was predicted and in a survey which was conducted where the users were made to judge between the recommendations from both the other systems.

Yu Zhu [1] proposes a design that selects user selection criteria based on items' attributes and users' rating history, and combine the criteria in an optimization framework for selecting users. By exploiting the feedback ratings, users' previous ratings and items' attributes, accurate rating predictions for the other unselected users are generated.

Muyeed Ahmed [7] proposes a clustering model to separate dissimilar users taken from MovieLens dataset and the results are compared in the three methods. Principal Component Analysis is used to make less the dimension for a improved clustering result. It is implemented such that none of the user data is lost after the process of clustering. Then, the rating which is present in the final column as result column is then used for the neural network.

III. RESEARCH METHODOLOGY

3.1Population and Sample

This dataset movie lens -20M describes five scale rating and free-text tagging features from MovieLens, a movie recommendation webservice. It contains 2 Million ratings and 465564 tag applications across 27278 movies. The data were created by 138493 users between January 09, 1995 and March 31, 2015. This dataset was generated on March 31, 2015, and updated on October 17, 2016.

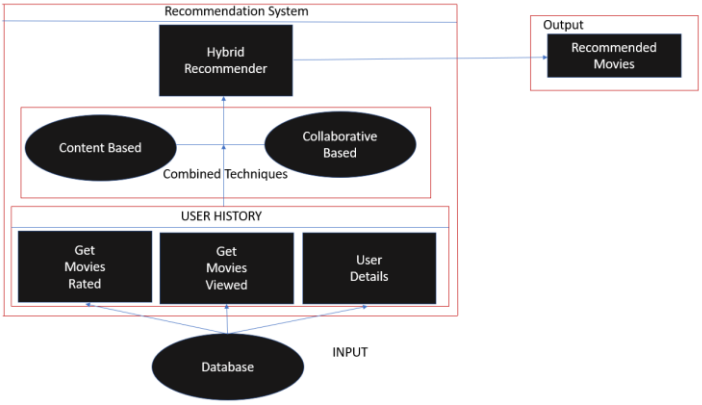
3.2 Data and Sources of Data

Random users were taken and they had rated 20 movies with no demographic information included. Here users are represented using user ids and the dataset contains about six files, genome-tags.csv, genome-score.csv, movies.csv, links.csv, tags.csv and ratings.csv.

3.3 Theoretical framework

User ids are the primary variables in the system. Movies with a minimum of one rating are included in the dataset and every user is given a unique user id rates the movies and the movies have movie ids. Ratings are made on a scale of 1-5, with increments of half star. Movie information is contained in file movies.csv.

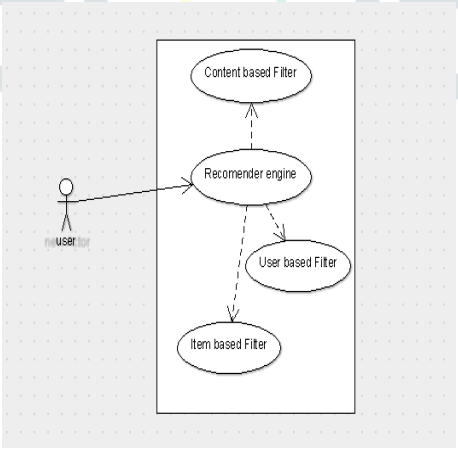
The proposed system uses hybrid approach that unifies both Content based and Collaborative based approaches so that the user can benefit the outcomes of both. We use two Collaborative approaches user based and item based to improve the accuracy of the recommendation system. The user based Collaborative filter is implemented based on Pearson product method correlation coefficient algorithm and the item based Collaborative filter is implemented based on Bayesian personalized ranking.



Architecture diagram

3.4 Tools and models

Content based and Collaborative based approaches combined together as hybrid filtering technique is used as the model, which employees item based and user based techniques for collaborative method.



Use case diagram

3.4.1 CONTENT BASED FILTERING

Content based filtering, recommends movies on the basis of a comparison between the description of the movie and user preferences. The content of each movie is represented as a set of descriptors or genres, typically the genre of the movie. The user content is presented with the same terms and built upon by analyzing the content of movies which the user had seen. Content based recommendations depend upon the features of movies, it is likely to be highly relevant to the interest of the user. This makes content based filtering very much valuable for places with huge libraries of a single type of content. Compared to the advanced math

involved in building a collaborative recommending system, the science behind the content based system is relatively straightforward. Content-based recommenders don't require other users to interact with any of the movies before it starts recommending it. In order to work accurately, a collaborative filter needs ratings, and not all users rate the movies constantly and that forms the primary dataset. In content based filter the user's preference towards the genres are obtained and recommendations are made based on the responses. The preference scale is from one to five, the preferences are obtained for each of the 18 genres. At first the genres most liked to least are sorted and then movies matching the preferred genres are found.

3.4.2 Genre Based Preference

User's preference towards the genres are obtained and recommendations are made based on the responses. The preference scale is from one to five, the preferences are obtained for each of the 18 genres. At first the genres most liked to least are sorted and then movies matching the preferred genres are found. Direct response from a user, normally in the form of a preference towards a genre, can be used to initialize higher or lower points on the significance of certain attributes and suggest movies according to the preferred genre.

3.4.3 COLLABORATIVE BASED FILTERING

Collaborative filtering filters the movies by using the previously made reviews of other users. It relies on the theory that the people who reviewed certain movies in the past might agree to it in the future. Suppose a person who wishes to see a movie, will ask for recommendations from friends and his peers. The recommendations given by some of the friends who have similar likings are heard by the person more than suggestions from other sources. This idea is used in the decision point on what movies to see.

They are based on collecting and analyzing a huge amount of data on the preferences of users and figuring out what the users might tend to like based on their resemblance to other users. Collaborative filtering approach does not depend on analyzable content and so it can precisely recommend complex movies without requiring an understanding of the movie itself.

3.4.3.1 TYPES OF COLLABORATIVE FILTERING

There are two methods in collaborative filtering which are user based and item based respectively. User based measures the similarity between the main user and other users. Item based measures the similarity between the movies that target users' ratings and interactions with and other similar movies. The core idea behind Collaborative filtering is that similar users have similar interest and that similar movies are liked by a user.

3.4.3.1 User Based Collaborative Filtering

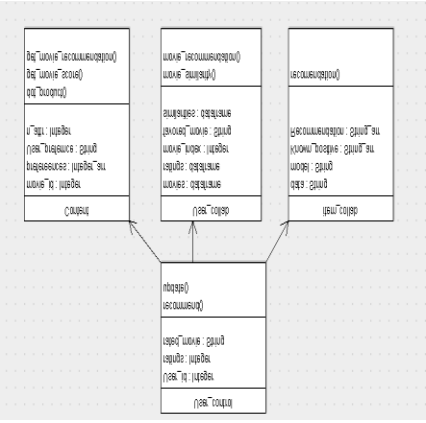
The user based collaborative filter works on Pearson product method correlation coefficient. It is the measure of linear correlation between two variables. The range of the values is between -1 and +1, where +1 denotes a total positive linear correlation, -1 denotes a total negative correlation and 0 denotes no correlation between the variables. We use the coefficients that have a value between 0.5 to 1.0 as a similarity and recommend them to the users.

3.4.3.2 Item Based Collaborative Filtering

The item based Collaborative filter works on the Bayesian Personalized Ranking algorithm. The algorithm coverts the filtering into a ranking problem by assuming that the users tend to like movies similar to that they have already liked. It does not use a virtual rating instead rates the candidate movies for the user. This algorithm uses a pair-wise interpretation of the implicit feedback matrix and reconstructs them for each user parts, that is a user's positive feedback signifies a preference of the user over an item that the user did not give any feedback on. A positive only response will be transformed into positive and negative feedback in terms of pairs of items, where the user prefers first over the second and correspondingly rephrased dislikes second over the first.

3.4.3.2 Comparison between those two Methods/Filtering techniques

In content based filtering results tend to be highly relevant and are transparent. Users can get started quickly and they get recommendations on new movie immediately and this method is more technically easier to implement. In collaborative based filtering benefits from large user bases and flexible across different domains, produces more serendipitous recommendations**.** It can capture more nuance around item.



Class diagram

IV. RESULTS AND DISCUSSION

4.1 Results

The proposed hybrid recommendation system containing the content and two Collaborative approaches works on the ml-20M data ser it is a five star rating from a movie recommendation webservice. The dataset contains 2 million ratings and 465564 tag applications across 27278 movies. These data were created by 138493 users. To provide content based recommendations, we used user preferences of different genres. Then the system was evaluated and the execution time of the algorithm was clocked as 00:04.401722 and 0:00:17.968693.

IV. ACKNOWLEDGMENT

This project was guided by Ms.T.Malathi , SRM INSTITUTE OF SCIENCE AND TECHNOLOGY.

REFERENCES

[1] Yu Zhu, Jinhao Lin, Shibi He, Beidou Wang, Ziyu Guan, 2018 **,** Addressing the Item Cold-start Problem by Attribute-driven Active Learning,IEEE Transactions on Knowledge and Data Engineering

[2] Bagher Rahimpour Cami, Hamid Hassanpour, Hoda Mashayekhi,2017,A Content-based Movie Recommender System based on Temporal User Preferences, 3rd Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)

[3] Muyeed Ahmed, Mir Tahsin Imtiaz, Raiyan Khan, 2018, Movie Recommendation System Using Clustering and Pattern Recognition Network

[4] *Ningning Yi, Chunfang Li, Xin Feng, Minyong Shi,2017,* Design and Implementation of Movie Recommender System Based on Graph Database, 14th Web Information Systems and Applications Conference

[5] Jeffrey Lund, Jeffrey Lund,2018, Movie Recommendations Using the Deep Learning Approach, IEEE International Conference on Information Reuse and Integration for Data Science

[6] Kunal Shah, Akshaykumar Salunke, Saurabh Dongare, Kisandas Antala,2017, Recommender Systems: An overview of different approaches to recommendations, National Conference on Innovations in information Embedded and Communication Systems (ICIIECS)

[7] Yeo Chan Yoon, Yeo Chan Yoon,2018, Movie Recommendation using Metadata based Word2Vec Algorithm, International Conference on Platform Technology and Service

[8] Shreya Agrawal, Pooja Jain,2017, An Improved Approach for Movie Recommendation System, International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)