

Content Based, Collaborative and Statistics Based Filtering Techniques in Recommendation System

Aarepu Lakshman¹, Dr. Yogesh Kumar Sharma², Dr. B.M.G. Prasad³

Research Scholar, Shri Jagdish Prasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, India¹

Associate Professor & HOD Dept of CSE & IT, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, India²

Professor, Dept of CSE and Dean PG Courses at Holy Mary Institute of Technology and Science Hyderabad, Telangana, India³

ABSTRACT: Finding information on a large web site can be a difficult and time consuming process. Recommender systems can help users to find information by providing them with personalized suggestions. Instead of providing a static experience in which users search for products, recommender systems help to increase interaction to provide a richer experience. Recommender systems can easily identify the recommendations autonomously for individual users based on past purchases and searches, and on other users behavior. Moreover, the user access histories are analyzed to derive the profiles of user interests and behaviors for user grouping. The content based, collaborative, and statistics-based filtering recommendation methods are proposed in this paper based on the favorite degrees of the users to the item groups, and the user groups they belong to. The result analysis of the proposed recommendation system is done on data of Movie Lens Website and it gives the precession of recommendation system using a single feature and set of features respectively for the three proposed recommendation methods of CB, COL and STA.

KEY WORDS: Recommender systems, Content based Filtering, recommendation methods and user profiles.

I. INTRODUCTION

The web is the place where billions to millions information sources are available as per user need. As there are huge amount of datas are available over web, the problem of information overload appears. Because of Information overload the mining results can take more time to search the required information or even can provide wrong prediction as sparsity of data. The rapid development of internet brings the problem of information overload: user receives too much potential valuable information to approach the most useful part, which causes the reducing of efficiency. Recently, many applications of network (web portal, search engine, data index, etc.) help users filter the information. However, all of these methods just satisfy the basic needs and do not consider the personalized services to solve the information overload problem.

Recommendation systems have become extremely common in recent years. Almost all of the large e-business systems have applied recommendation system in different ways, such as Amazon, eBay, etc. The recommendation system are the supporting system which is used to help users to find information services, or products such as Books, Music, Movie, Digital Products, Web sites and TV Programs by analyzing the suggestions from other users [4]. Based on the past history the recommender system provide list of items by predicting which item are most suitable to user, preferences and constraints. Recommendation systems use different algorithm and methods to provide personalized recommendations. Recommendation system is also called information filtering system, or recommendation engine used to recommend informational items.

As an important method of information filter, recommendation system provides personalized services to users through identifying and predicting user preferences and become a most useful method helping user to deal with information overload [5]. Recommender system assist and augment this natural social process to help people to sieve through available books, articles, web pages, electronic products, grocery items and so forth to find the most interesting and valuable information for users. The idea of content based filtering is that users are interested in items that are similar to item the users previously liked.

II. LITERATURE REVIEW

Recommender systems generate personalized preferences using information filtering techniques and algorithms with a goal to support decision making of the users. Recommender systems are mainly used in various domains like online booking, online shopping, audio and video recommendations and so on. The concept of recommendation is not new but in use from many years, the difference is due to more number of users asking for recommendations among thousands to millions of choices. It has become a tedious job to recommend someone appropriately without filtering the data for relevant choices. It depends on the factors like users rating given to collection of items based upon the user satisfaction level, their likes and dislikes, age group, gender, occupation, region or locality, community etc. There are various problems and challenges consequently affect the performance of recommender systems. The main problem is one of the challenges and that arrives when either a new user enters into a system or a new product arrives in catalogue. Both situations lead to difficulty in predicting user preference in the absence of availability of sufficient user rating history [3].

However, the system in Kuo and Shan (2002) cannot actually reflect this preference for the users during recommendation. Moreover, due to the inherence of the content based filtering approach, this system cannot provide any surprising recommendation results as in the collaborative approach (6). Ringo in Sharadanand and Maes (1995) is a pioneer collaborative music recommendation system [7]. In Ringo, each user is requested to make ratings for some music objects. These ratings constitute the personal profile. Several algorithms are proposed to measure the similarity between two user profiles. For collaborative recommendation, only the ratings of the users whose profiles are similar to the target user are considered. Whether a music object will be recommended is then based on the weighted average of the ratings considered.

For example, the user may prefer the music objects with slower tempo.

However, recommendation systems adopting the content based filtering approach can only recommend the data items in which the user has indicated his/her inter standard. Other potential interesting data items of the user cannot be explored in such recommendation systems if he/she has never accessed before. Instead of computing the similarities between the features of the data items and the user profiles, the collaborative approach computes the similarities between the user profiles. Users of similar profiles are grouped together to share the information in their profiles. The main goal of the collaborative approach is to make the recommendation among the users in the same group.

III. PROPOSED SYSTEM

The System Block Diagram as shown in figure (1) consists of six function blocks, namely, contains the Feature Extraction, the classifier, the profile manager, the recommendation module, the user interface and the database. In this system the recommendation process first starts by gathering information about the items like author, title, cost etc and uses feature extraction and information indexing.

In this system the feature extraction and representation could be achieved automatically like news from papers but human editors which have to manually insert features from items like movies and songs. Recommender systems help to match users with items and for this various recommender systems have been designed according to availability of exploitable data, implicit and explicit user feedback, domain characteristics etc. Recommender Systems are classified according to approach/paradigm used for predicting preferences. Using the features extracted, a dynamic classifier is designed for items grouping. Each incoming item is properly assigned to certain similar item group by the classifier.

Organized by

Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

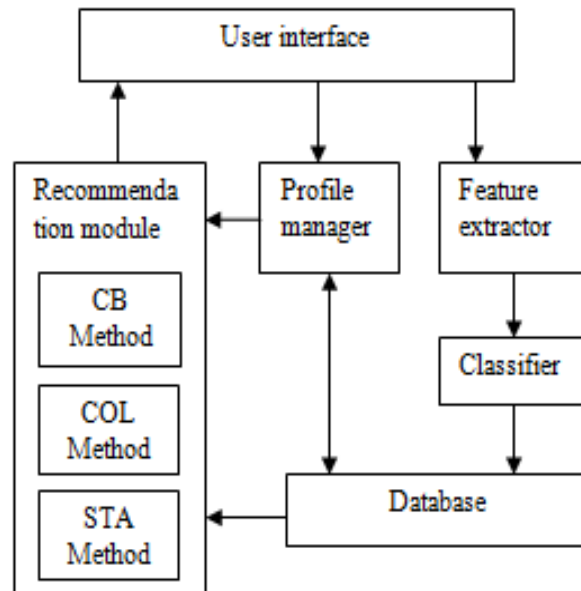


Fig. 1: PROPOSED RECOMMENDATION SYSTEM

The profile manager is implemented for the purpose of updating the access histories of the users. When the user accesses a item from the list of item objects or the recommendation results, the profile manager records the information of the item object in the user’s access history. Using the information recorded by the profile manager, three recommendation mechanisms, i.e., the content-based (CB) method, the collaborative (COL) method, and the statistics-based (STA) method, are designed for different users’ needs.

A. content-based (CB) method

This system uses content based (CB) filtering for processing information and data from various sources and try to extract useful features and element about the contents of the items and the constraint based filtering uses features of items to determine their relevance. Content based methods provide the recommendations by analyzing the description of the items that have been rated by the user and the description of items to be recommended. More number of algorithms has been proposed for analyzing the content of text documents and finding similarities in this content that can serve as the basis for making recommendations. Main goal of classification learners is to learn a function that predicts which class a document belongs to. Other algorithms would use a regression problem in which the goal is to learn a function that predicts a numeric value of the rating of the document. There are two important sub problems in designing a content based filtering system. The first is finding a representation of documents and next is to recommend for unseen documents.

As described above, the recommendation of the CB method depends on the users interests and the interests are derived from the users access histories. Therefore, the users will never get a recommendation of the item objects belonging to the similar item groups they never accessed before. That is, the CB method tends to provide expected and interesting music objects for users. Based on the collaborative approach, the purpose of the COL method is to provide surprising findings due to the information sharing between relevant users. In recommendation, collaborative filtering arrives to model the prior user behaviour and this model can be constructed from a single user’s behaviour or also from the other user behaviour that have similar traits.

When it takes other users behaviour into account, collaborative filtering uses group knowledge to form a recommendation based on like users. The recommendations are made from automatic collaboration of multiple users and filtered on those who exhibit similar preferences or behaviors. By using the information from many user who

Organized by

Department of ECE, Adhityamaan College of Engineering, Hosur, Tamilnadu, India

subscribe to and read blogs, you can group those users based on their preferences. From this information, you identify the most popular blogs that are read by that group.

C. The STA method

The third recommendation method is based on the statistics. We define the long-term hot music group as the music group containing the most music objects in the access histories of all users. Similarly, we define the short-term hot music group as the music group containing the most music objects in the latest five transactions in the access histories of all users. When the user chooses this recommendation method, the MRS recommends the latest N music objects (which have not been accessed by the user), half from the long-term hot music group and the other half from the short-term hot music group to the user.

IV. RESULTS

The dataset has downloaded from MovieLens Website. It contains 100004 ratings and 1296 tag applications across 9125 movies. These data were created by 671 users between January 09, 1995 and October 16, 2016. Users were selected at random for inclusion. All selected users had rated at least 20 movies.

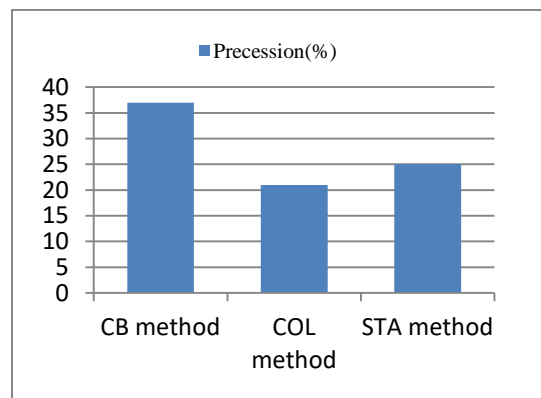


Fig. 2: THE PRECISION OF RECOMMENDATION SYSTEM BY SINGLE FEATURE

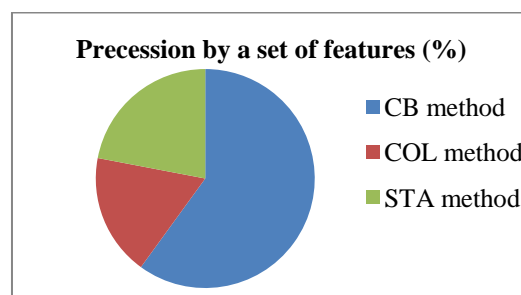


Fig. 3: THE PRECISION OF RECOMMENDATION SYSTEM BY FEATURE SETS

For front-end Java Jdk-7 with Eclipse Neon and Backend Microsoft SQL Server 2012 is used. To provide Graphical User Interface Tomcat Server is used. After Login User has Two toggle buttons: 1) User Playlist 2) Follower 3) Content - based recommendation 4) Collaborative recommendation 5) statistics method recommendation.

Recommendation is measured by *precision* defined as follows:

$$Precision = \frac{N_A}{N} \text{----- (1)}$$

Where N is the number of music objects in the recommendation list and NA is the number of music objects which the user accesses in the recommendation list. The precession results are shown in figure (2) and figure (2) for a single feature and set of features respectively. Note that we select 20 as N in the experiments.

V. CONCLUSION

In World Wide Web, the overload of information leads to the necessity of recommender systems to generate efficient solutions has evolved. Nowadays finding the right recommender for evaluating the reliability of recommender systems is an essential feature. For each user, the user's favorite degrees to the required item groups are determined in the proposed CB method. For providing surprising item objects, further take into account the opinions from the users of the same user group. The COL method is then proposed for this purpose. This classifier avoids the overhead from the task of manual classification. Moreover, we derive both the interest and behavior profiles from the users access histories for user grouping. The proposed technique to derive user profiles has the adaptability to large number of accessed music objects. We also perform a series of experiments to show that our recommendation system is practical.

REFERENCES

- [1] L Feng and G. Wei-Wei, "Recommendation Algorithm Based on Tag Time Weighting [C]", 2018 International Conference on Smart City and Systems Engineering (ICSCSE), 2018
- [2] Natrah Abdullah, Nur Amirah Mustapar and Nor Laila Md Noor, "A review towards developing a moment of information overload model", 2016 4th International Conference on User Science and Engineering (i-USER), IEEE, Year: 2016.
- [3] Hongwu Ye, "A Personalized Collaborative Filtering Recommendation Using Association Rules Mining and Self-Organizing Map", International Journal of Computer Science & Information Technology (IJCSIT) Vol 6, No 2, April 2014 DOI:10.5121/ijcsit.2014.6204
- [4] A. Bellogí I. Cantador, F. Dí et al., "An empirical comparison of social, collaborative filtering, and hybrid recommenders," ACM Trans. On Intelligent Systems and Technology, vol. 4, no. 1, pp. 1-37, January 2013
- [5] G. Adomavicius and A. Tuzhilin. "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Transactions on Knowledge and Data Engineering, vol. 17, Jun. 2005, pp. 734-749.
- [6] Kuo, F.F. and Shan, M.K., "A Personalized Music Filtering System Based on Melody Style Classification," In Proceedings of IEEE international Conference on Data Mining, 2002, pp. 649-652.
- [7] Sharadanand, U. and Maes, P., "Social Information Filtering: Algorithms for Automating 'Word of Mouth'," In Proceedings of CHI'95 Conference on Human Factors in Computing Systems, 1995, pp. 210-217.
- [8] Study of Recommender System - www.ganna.com ! My First Blog - <https://tejaswinidevappa.wordpress.com/2015/01/20/study-of-recommender-system-www-gaana-com/>
- [9] Music for everyone - Spotify www.spotify.com
- [10] Frederik Pruifer, Music Recommendation at Spotify - How Spotify Recommends music.