A Survey on Algorithms Used in Recommender Systems

Venkatesh P

PG Student, Dept. of Studies in CS&E., UBDTCE Davanagere, Karnataka, India

ABSTRACT: Recommender system can be defined as the subclass of information filtering system which attempts to give the guidance to the users regarding the useful services based on their personalized preferences, past behavior or based on their similar likings with other users. The various approaches of recommendation systems, like content-based, collaborative filtering, hybrid, etc, can further be classified according to their algorithmic technique as memory-based (heuristic) or model-based recommendation algorithms. Service recommender systems provides appropriate recommendations of services like movies, hotels, gadgets, etc, leading to an increase in the amount of data on the web, known as Big Data. It is becoming difficult to capture, store, manage and analyze such big data that affects the service recommender systems with issues like scalability and inefficiency. Also many existing service recommender system provides the same recommendations to different users based on ratings and rankings only, without considering the taste and preference of an individual user. This paper presents a survey on various recommendation algorithms, elaborating all its types along with its drawbacks. The paper also focuses on the solutions to overcome these drawbacks and provide apt recommendations to the users. It also deals with the solution to provide apt recommendations of the services to the users in big data environment. The issues of scalability and inefficiency while managing big data can be solved by using a distributed computing platform known as Hadoop.

KEYWORDS: Service recommender system, Recommendation algorithms, Big data, Distributed computing platform,Hadoop

I. INTRODUCTION

Information filtering system is a system that removes dispensable or unnecessary information beforerepresenting it to the human user. It is a way of sifting through the overabundance of data on the web. These information filtering system have a subclass called recommender systems or recommendation systems that predict the 'rating' or 'preference' that user would give to an item. Service recommender systems provide appropriate recommendations of services to the users. These service recommender systems have become popular in variety of practical applications like recommending the users about movies, music, hotels, news, books, research articles, search queries, social tags, and products in general. However, there are also recommender systems for experts, jokes, restaurants, financial services, life insurance, and twitter followers, etc. But most of the existing recommender systems provide recommendation to the active/present user based on the ratings and rankings provided by the past users. Moreover with the growing number of alternative services, recommending the services effectively, that the users prefer has become an important research issue. The service recommender systems present the same ratings and rankings of the services to the different users and also provide the same recommendations to them without considering the users' personal likings and taste. Therefore these systems fail to meet users personalized requirements.

Nowadays data have also become a torrent flowing into every area of the global economy. It has now swept into every industry and business function. Companies churn out a large volume of transactional data, by capturing information about their customers, suppliers, and operations. Also the online users, consumer devices including PCs and laptops, various new online applications, networked sensors embedded devices like automobiles and industrial machines, smart-phones, social media sites, etc, have increased explosively, thus increasing the amount of data on web. Such large amount of data is called as Big-data. Big-data refers to a dataset, immense in nature and difficult to capture, store, manage, process and analyze with the available current technology within the tolerable speed and time. The growing bulk of multimedia content has played a major role in the exponential growth in the amount of big data. This Big-data management is a serious problem to all the companies and IT industries and also poses a heavy impact on service

recommender systems. Issues like scalability and inefficiency which comes along while managing and analyzing large amount of dataset, are of great problem for the service recommender systems.

II. AN OVERVIEW OF RECOMMENDATION ALGORITHM

The various recommendation methods can be classified into various types based on the information orknowledge source they use for making the apt recommendations. The diagrammatic representation of the classification is shown in the following figure 2.1. They are classified as [1]:

- Content-based Recommendations
- Collaborative Recommendations
- Knowledge-based Recommendations
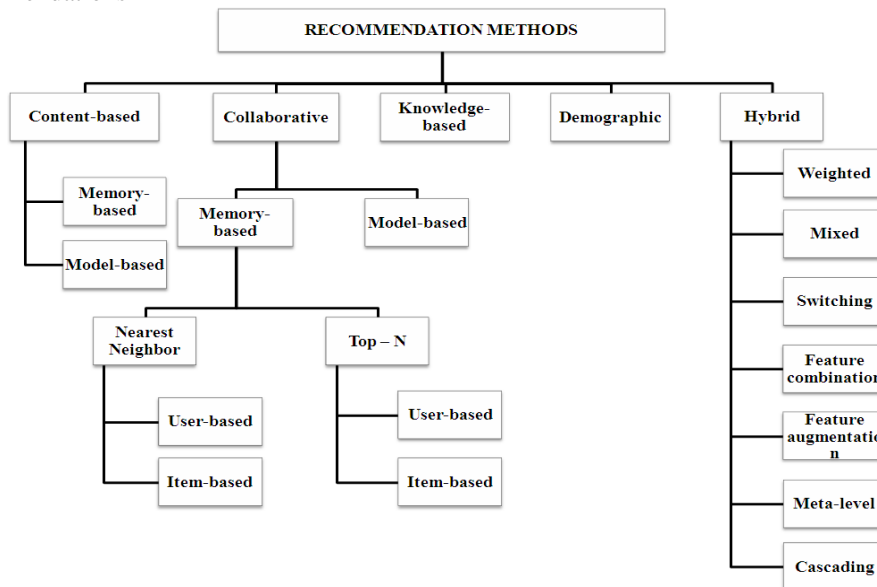- Demographic Recommendations
- Hybrid Recommendations

Fig.1. Classificationof recommendationmethods

1.1 Content-based Recommendations:

In content-based recommendation, the items recommended to the user are similar to those items that heor she preferred in the past. For example, in a book recommendation application, to recommend the books to the user u, the content-based recommendation system looks for the similarities among the books the user u has rated in the past. It uses a technique to analyze a set of documents and descriptions of items previously rated by the user. It then builds a profile or model of the users interests based on the features of those rated items by using keyword analysis techniques from information retrieval. Then using the user profile, the recommender system can filter out the suggestions that would best fit for the user [2]. It can either be:

- Memory/Heuristic based - uses frequency, inverse document frequency (TF-IDF) text retrieval method
- Model based - uses Decision trees, neural networks, Bayesian classifiers, Clustering or vector-based representation.

There are various limitations faced by it such as ramp-up problem i.e. new itemscannot be recommended to any user until they get some sort of rating. These recommendation systems also faces the plasticity problem i.e. once a user's profile has been set up in the system, it is hard for the system to adapt to changes in the user's preferences [3].

2.2 Collaborative Recommendations:

In collaborative filtering recommendation system the services recommended to the user are those thatusers with similar tastes preferred in the past i.e. they receive recommendations based on the people who have similar tastes and

preferences. The basic assumption in CF is that user A and user B's personal tastes are correlated if both users rate n items similarly. To generate recommendation using collaborative filtering systems, first we need to gather the ratings from users and maintain user's ratings in a database. Then the correlations between the pairs of users to determine a user's neighbors in taste space is computed followed by computing the ratings of these neighbors to make recommendations [4]. Collaborative recommendations can be classified into following types:

- Item-based CF - The predicted rating depends on the ratings of other similar items by the same user.
- User-based CF - The prediction of the rating of an item for a user depends upon the ratings of the same item rated by similar users [5].
- Collaborative Filtering systems can be separated into 2 classes depending on the algorithmic technique as given below:
- Memory-based (Heuristic) Recommendation Systems - They make predictions by operating on data (users, items and ratings) stored in memory. They can be classified as :

(i) Nearest Neighbor algorithms - Users similar to the current user with respect to preferences are called as neighbors. This algorithm deals with discovering the past users who had historically the similar taste as that of the new user. Items that the neighbors like are then recommended to the new user, as he will probably also like them. User-based nearest neighbor algorithms generate predictions for a given user based on ratings provided by users in the neighborhood. Item-based nearest neighbor algorithms generate predictions based on similarities between items.

(ii) Top-N recommendation algorithms – It recommends a set of N top-ranked items that will be of interest to a certain user. The user-item matrix is analyzed to correlate different users or items and use them to compute therecommendations. User-based Top-N Collaborative Filtering Algorithms identify the k most similar users(nearest neighbors) to the active user. Item-Based Top-N Collaborative Filtering Algorithms compute the k most similar items for each item according to the similarities between them [6].

2.3 Knowledge-based Recommendations

Knowledge-based Recommender Systems suggest to users items of their interest, on the basis of someunderstanding of both items' characteristics and users' profiles. It utilizes the knowledge about users and products and reasons out what products meet the users' requirements. In order to properly work, this kind of recommender systems need a thorough modeling of items and users; the usual barrier to their development is, therefore, the availability of the necessary knowledge and its maintenance over time. The advantage with this is that it doesn't have the "ramp-up" problem since its recommendations don't depend on any database of user ratings. Moreover they need a minimal amount of users, i.e. they do not require a huge amount of data to compute recommendations, differently from other classes of recommender system. The drawback with this system is that it requires an engineered knowledge database to make useful recommendations. This knowledge base has to be updated to keep up with the ever-changing consumer ratings and preferences [7].

2.4 Demographic Recommendations

Demographical data such as age, gender, social class, education, location, etc. can be used and obtainedexplicitly and implicitly. Based on the demographical data a user profile can be created. With a demographic profile, certain matching items can be recommended. For instance teenagers prefer different products than the elderly and the rich may want different products than middle and lower classes and are willing to pay more. Therefore demographic recommendation categorizes the user based on personal attributes and makes recommendations based on demographic classes. The advantage of a demographic approach is that the user-item ratings are not used, so new users can get recommendations before they have rated any item. And knowledge about the items and their features is not needed; therefore the technique is domain independent. The disadvantage of the demographic approach is that gathering the required demographic data leads to privacy issues. Demographic classification is also too crude for highly personalized recommendations [8].

2.5 Hybrid Recommendations

Many of the existing recommender systems utilize only a handful of basic recommendation techniqueslike content-based, collaborative, demographic, utility-based and knowledge-based. These techniques when used individually have many complementary advantages and disadvantages. This fact has provided a stimulate for research

in hybrid recommender systems, that combine various techniques for improved performance. Various hybrids have been explored, including the hybridization techniques namely weighted, mixed, switching, feature combination, feature augmentation, and meta-level. They are described as follows [9]:

- Weighted: Implementing Collaborative and Content-based methods separately and then combining their predictions.
- Switching: A certain switching criterion is used by the system to interchange between two recommendation systems operating on the same object.
- Feature Combination: Features from different recommendation systems' data sources are put into a single recommendation algorithm.
- Cascading: For this category, one recommendation system refines the results given by another.
- Meta Level: In this case, a feature such as a model learned by one recommendation is used as input to another.
- Feature Augmentation: The output of one system is used as an input feature to another; for example, using the model generated by one to generate features that are used by another.
- Mixed: Incorporates two or more techniques at the same time; for example: Content-based and Collaborative Filtering.

III. CONCLUSION

In this paper various methodology have been discussed which are used to generate recommendationsfor the active users on the web applications. Various pros and cons of various methodology are stated which gives rise to the need of using hybrid methodology for recommendations. Also adopting various combinations of recommendation methods helps to increase the accuracy in recommending the appropriate choices of items to the users. Hadoop, a distributed cloud computing platform is one of the main approaches used to manage large amount of data in these service recommendation systems. To achieve scalability in service recommendation based on keyword aware and also to solve inefficiency in handling the large amount of services during recommendation can be the scope of further research in this area.

REFERENCES

[1]. Recommendation system: http://en.citizendium.org/wiki/Recommendation_system
[2]. Michael J. Pazzani and Daniel Billsus, "Content-based Recommendation Systems".
[3]. Content based Recommendations: http://findoutyourfavorite.blogspot.in/
[4]. Collaborative Filtering: http://en.citizendium.org/wiki/Collaborative_filtering
[5]. SongJie Gong, "A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering," journal of software, vol. 5, no. 7, july 2010
[6]. BadrulSarwar, George Karypis, Joseph Konstan, and John Ried, "Item-based Collaborative Filtering Recommendation Algorithms," May 1-5, 2001, Hong Kong
[7]. Daniele Dell'Aglio, Irene Celino, and Dario Cerizza, "Anatomy of a Semantic Web-enabled Knowledge-based Recommender System".
[8]. ManishaHiralall, "Recommender systems for e-shops," VrijeUniversiteit, Amsterdam, 2011
[9]. R. Burke, "Hybrid Recommender Systems: Survey and Experiments," User Modeling and User-Adapted Interaction, Vol. 12, No.4, pp. 331-370, 2002
[10]. G. Adomavicius, and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of- the-Art and PossibleExtensions," IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.6 pp. 734-749, 2005.
[11]. Z. D. Zhao, and M. S. Shang, "User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop," In the third InternationalWorkshop on Knowledge Discovery and Data Mining, pp. 478-481, 2010.

BIOGRAPHY

Venkatesh P is a PG Student in the Department of studies in Computer Science and Engineering, UBDTCE Davanagere, Visvesvaraya Technological University, Belagavi. He received Bachelor degree in 2014 from BIET, Davanagere . His research interests are Big Data, web 2.0, Enterprise mobilization etc.