A Survey of Recommendation System: Research Challenges

Lokendra Singh

A Survey of Recommendation System: Research Challenges

Lalita Sharma[#1], Anju Gera[#2]

[#1]M-Tech Scholar & Department of Computer Engineering, M.D. University, BSAITM,Faridabad, India
[#2]Assistant Professor & Department of Computer Engineering, M.D. University,BSAITM, Faridabad, India

***Abstract*— Recommendation is a process which plays an important role in many applications as WWW.The main objective of this paper is to show various challenges regarding to the techniques that are being used for generating recommendations. Recommendation techniques can be classified in to three major categories: Collaborative Filtering, Content Based and Hybrid Recommendations. By giving the overview of these problems we can improve the quality of recommendations by inventing new approaches and methods, which can be used as a highway for research and practice in this area.**

Keywords:-Collaborative Filtering, Content-Based Recommendation, Recommendation System, Sparsity Problem.. Cold Start,over specialization.

Introduction

Recommender systems or recommendation systems are a subclass of information filtering system that seek to predict 'rating' or 'preference' that a user would give to an item (such as music, books or movies) or social element (e.g. people or group) they had not yet considered, using a model built from the characteristics of an item (content based approaches) or the user's social environment (collaborative filtering approaches). Although many different approaches to recommender systems have been developed in the past few years, the interest in this area still remains high due to growing demand on practical applications, which are able to provide personalized recommendations and deal with information overload. These growing demands pose some key challenges to recommender systems and to deal with these problems many advanced techniques are proposed like content boosted collaborative filtering, clustering based filtering, combining item based and user based similarity and many more. Despite of these advances, recommender systems still require improvement and thus becoming a rich research area.
In this paper, before discussing the major limitations of recommendation methods, the comprehensive survey of recommendation approaches is provided. The discussion of various approaches and their limitations in a proper manner thereby provides the future research possibilities in recommendation systems.

I.

RECOMMENDATION SYSTEM: GENERAL CONCEPTS

Recommendation System is an intelligent system that makes suggestion about items to users that might interest them. Some of the practical applications that use such systems include recommending books, cd etc. on Amazon.com, movies by Movielens, music by last.fm and news at VERSIFI technologies.

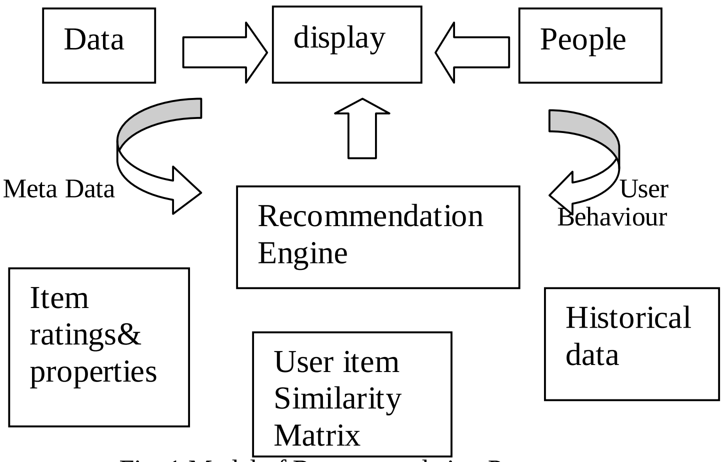


Fig. 1 Model of Recommendation Process

In a recommendation-system application there are two classes of entities, which we shall refer to as users and items. The formal definition of recommender system is:

- *C*: The set of all users
- *S*: The set of all possible items that can be recommended, such as books, movies, or restaurants.
- U: A utility function that measures usefulness of a specific item $s \in S$ to user $c \in C$, i.e., $U: C\times S \rightarrow R$, where R is a totally ordered set.

The space *S* of possible items can be very large, ranging in hundreds of thousands or even millions of items in some applications, such as recommending books or CDs. Similarly, the user space can also be very large—millions in some cases. For each user $c \in C$, we want to choose such item $s' \in S$ that maximizes the user's utility is desired.
More formally:

$$\forall c\ \varepsilon\ C, Sc = \max_{s\ \varepsilon\ S} u(C,S)$$

In recommender systems the utility of an item is usually represented by a *rating*, which indicates how a particular user liked a particular item. In general, rating is done on scale, for instance, if movies are rated on a scale of 1 to 5, then a movie rated 5 by a user means it is highly liked by user while 1 rating denotes dislike. Further, each element of user space U can be defined with a profile that includes user characteristics like userid, age, gender, occupation etc. and each element of item space S can be defined using item characteristics. For example, in a movie recommender system, a movie can be defined by its id, genre, release date, director, actors etc. Usually, rating is not done on a complete dataset or space C×S and thus only rating on subset is available. The main aim of a recommender system is to predict ratings of the non-rated user/item combination and thus providing appropriate recommendations. A recommender system may either provide the highest estimated rating item or alternatively provide a list of top N items as recommendation to a user or set of users.

II. RECOMMENDATION SYSTEM TECHNIQUES

Recommendation systems are usually classified on the basis of rating estimation technique:

A. *Collabrative Filtering Process*

The Collaborative filtering (CF) systems work by collecting user feedback in the form of ratings for items in a given domain and exploiting similarities in rating behaviour amongst several users in determining how to recommend an item. CF systems recommend an item to a user based on opinions of other users. For example, in a movie recommendation application, CF system tries to find other like-minded users and then recommends the movies that are most liked by them.

The task of traditional collaborative filtering recommender algorithm concerns the prediction of the target user's rating

Item User	Item1	Item2		Item n	Item j
User1					
User 2					
User i					
User m					

Input Rating Matrix CF Algorithm

Predictions Rij
Recommendations

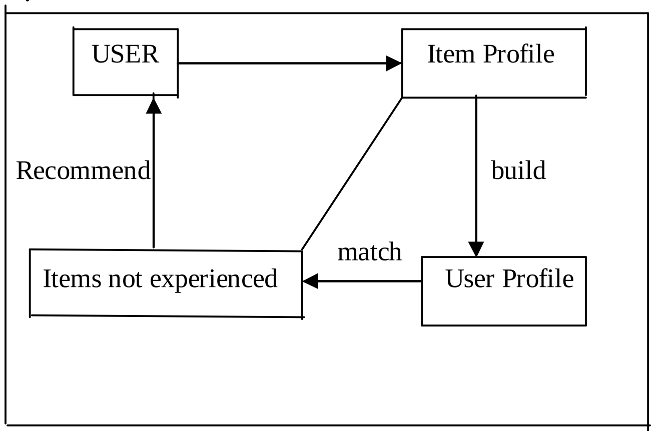
for the target item that the user has not given the rating, based on the users' ratings on observed items.

- CF algorithms represent the entire user-item space as a rating matrix 'R'. Each entry Rij in matrix represents the preference score (rating) if the ith user on the jth item. Each individual rating is within a numerical scale and it can be 0 as well, indicating that user has not yet rated this item.
- CF problem includes the estimation or prediction of rating for the yet unrated item. For the prediction of rating, similarities between items and users are calculated using different approaches. Thus, the two related problems consist in finding set of K users that are most similar to a given user and finding set of K items that are most similar to a given item.
- Finally using these similarities, recommendations that are produced at output interface can be of two types: Prediction and Recommendation.
- Prediction is a numerical value, Rij, expressing the predicted score of item j for the user i. The predicted value is within the same scale that is used by all users for rating.

B. *Content-Based Process*

Content based recommendation systems recommend an item to a user based upon a description of the item and a profile of the user's interests. Such systems are used in recommending web pages, TV programs and news articles etc. All content based recommender systems has few things in common like means for description of items, user profiles and techniques to compare profile to items to identify what is the most suitable recommendation for a particular user.

In content-based recommendation methods, the utility $u(c, s)$ of item s for user c is estimated based on the utilities $u(c, s_i)$ assigned by user c to items $s_i \in S$ that are "similar" to item s. Content-Based recommender systems make suggestions upon item features and user interest profiles. Typically, personalized profiles are created automatically through user feedback, and describe the type of items a person likes. In order to determine what items to recommend, collected user information is compared against content features of the items to examine . As shown below



- System has a huge database consisting of the items to be recommended and the features of these items and it is termed as Item Profile.
- The users provide some sort of information about their preferences to the system. Combining the item information with the user preferences, the system builds a profile of the users.
- According to the information existing in a target user's profile, the system recommends suitable items to the user.

One can make better personalized recommendation by utilizing the features of items and users.

An item profile is defined by its important features. For instance, a movie can be described by its title, genre, language, country, actors etc. Depending upon the weighing procedure, similarity between two items can be calculated. Depending on the domain, features can be represented either by Boolean values or by a set of restricted values. For example, imagine we want to analyse a set of newspaper articles about different kind of topics. While a Boolean value could indicate whether a word is contained in an article or not, an integer value could express the number of times a word appears.

To build a user profile, information of a user can be used. In MovieLens dataset, users are described using demographic information that includes age, gender, occupation and zipcode. User information might be provided explicitly by the individual person or gathered implicitly by a software agent.

- Explicit user information collection basically relies on personal input by the user. A common feedback technique is the one that allows users to express their opinions by selecting a value of range. However, filling out forms or clicking checkboxes places a burden on the user. Profiles might be imprecise, because the user is not willing to spend a lot of time providing personal information or the user information is already outdated.
- Implicit feedback does not require any additional intervention by the user during the process of constructing profiles. Moreover, it automatically updates as the user interacts with the system.

Thus, systems that collect implicit feedback are more likely to be used in practice

C. Hybrid Process

Hybrid recommenders are systems that combine multiple recommendations techniques together to achieve a synergy between them. Several researchers have attempted to combine collaborative filtering and content based approaches in order to smoothen their disadvantages and gain better performance while recommendations. Depending on domain and data characteristics, several hybridization techniques are possible to combine CF and CB techniques which may generate different outputs. Different ways of hybridization are:

- Implementing CF and CB separately and combine their predictions.
- Incorporating some content based characteristics into collaborative approach.
- Incorporating some collaborative characteristics into content based approach.
- Constructing a general unifying model that incorporates both content-based and collaborative characteristics.

Many hybrid approaches are based on CF but CB methods are used to maintain the user profiles and such profiles are used to find similar users.

III. RESEARCH CHALLENGES TO RECOMMENDATION SYSTEM:-

Various techniques used in a recommender system experiences some of the hurdles that may be described in terms of basic problems as:

A. Sparsity Problem

Sparsity problem is one of the major problems encountered by recommender system and data sparsity has great influence on the quality of recommendation. Generally, data of system like MovieLens is represented in form of user-item matrix populated by ratings given to movies and as no. of users and items increases the matrix dimensions and sparsity evolves. The main reason behind data sparsity is that most users do not rate most of the items and the available ratings are usually sparse. Collaborative filtering suffers from this problem because it is dependent over the rating matrix in most cases. Many researchers have attempted to reduce this problem; still this area demands more research.

B. Cold Start problem

Cold start problem refers to the situation when a new user or item just enters the system. Three kinds of cold start problems are: new user problem, new item problem and new system problem. In such cases, it is really very difficult to provide recommendation as in case of new user, there is very less information about user that is available and also for a new item, no ratings are usually available and thus collaborative filtering cannot make useful recommendations in case of new item as well as new user. However, content based methods can provide recommendation in case of new item as they do not depends on any previous rating information of other users to recommend the item.

C. Scalability

Scalability is the property of system indicates its ability to handle growing amount of information in a graceful manner. With enormous growth in information over internet, it is obvious that the recommender systems are having an explosion of data and thus it is a great challenge to handle with continuously growing demand. Some of the recommender system algorithms deal with the computations which increase with growing number of users and items. In CF computations grow exponentially and get expensive, sometimes leading to inaccurate results. Methods proposed for handling this scalability problem and speeding up recommendation formulation are based on approximation mechanisms. Even if they improve performance, most of the time they result in accuracy reduction.

D. Over Specialization Problem

Users are restricted to getting recommendations which resemble to those already known or defined in their profiles in some cases and it is termed as over specialization problem. It prevents user from discovering new items and other available options. However, diversity of recommendations is a desirable feature of all recommendation system. After solving the problem using genetic algorithms, user will be provided with a set of different and a wide range of alternatives.

IV CONCLUSION

The Several recommendation systems have been proposed that are based on collaborative filtering, content based filtering and hybrid recommendation methods but these have some problems which are the challenges for research work. It is required to work on this research area to explore and provide new methods that can reduce the challenges and provide recommendation in a wide range of applications while considering the quality and privacy aspects. Thus, the current recommendation system needs improvement for present and future requirements of better recommendation qualities.

ACKNOWLEDGMENT

I would like to express gratitude to my guide Mrs. Anju Gera for supporting me to do this.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEETKDE: IEEE Transactions on Knowledge and Data Engineering, 17, 2005.

[2] Prodan Andrei-Cristian, "Implementation of a Recommender System Using Collaborative Filtering", Studia univ. Babes_bolyai, Informatica, volume lv, number 4, 2010.

[3] Michael J. Pazzani and Daniel Billsus, "Content Based Recommendation System."

[4] Claypool, M., A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin, "Combining content-based and collaborative filters in an online newspaper". In ACM SIGIR'99, Workshop on Recommender Systems: Algorithms and Evaluation, August 1999.

[5] Debnath, Ganguly and Mitra, "Feature Weighting In Content Based Recommendation System Using Social Network Analysis".

[6] MovieLens Dataset, USA, 2003, University Minnesota.

[7] Balabanovic, M. and Y. Shoham. Fab, "Content-based, collaborative recommendation", Communications of the ACM, 40(3):66-72, 1997.

[8] Sarwar, Badrul M., George Karypis, Joseph A. Konstan and John T. Riedl, "Application of Dimensionality Reduction in Recommender System – A Case Study", In ACM WebKDD Workshop, 2000.

[9] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigentaste, "A constant time collaborative filtering algorithm", Information Retrieval, 4(2):133–151, 2001.

[10] Boddu Raja Sarath Kumarmaddali and Surendra Prasad Babuan:, "Implementation of Content Boosted Collaborative Filtering Algorithm", IJEST.

[11] J. Wang, A. P. de Vries, and M. J. T. Reinders, "Unifying User-Based and Item-Based Collaborative Filtering Approaches by Similarity Fusion". In SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, 2006.

[12] C. Desrosiers and G. Karypis, "Solving the Sparsity Problem: Collaborative Filtering via Indirect Similarities", Technical Report, Dec 2008.

[13] Terveen and Hill, "Beyond Recommendation System in HCI in the New Millennium, Addison-Wesley, 2001 p. 4 of 21.