

Recommender Systems: Types of Filtering Techniques

Iateilang Rynghsai

Department of Computer Science & Engineering and IT
Don Bosco College Of Engineering & Technology
Assam Don Bosco University
Guwahati, Assam-781017, India

L. Chameikho

Department of Computer Science & Engineering and IT
Don Bosco College Of Engineering & Technology
Assam Don Bosco University
Guwahati, Assam-781017, India

Abstract— At present, Recommender systems (RS) plays a very important role in supporting e-commerce stores and Online Social Networks (OSN) by suggesting useful recommendations to the users based on the use of networked databases. Different techniques for filtering are used by RS. The Collaborative Filtering (CF) technique filters or evaluates item through the opinions of other people. Demographic Filtering (DF) technique uses the demographic data of a user to determine which items may be appropriate for recommendation. Content-Based Filtering (CBF) technique recommends items for a user based on the description of formerly evaluated items and information obtainable from the content. This paper presents an overview of these basic filtering techniques which are being used in e-commerce stores and OSN to extract data from different users and then utilize the information for a better RS.

Keywords—Recommender systems (RS), E-commerce stores, Social networks, Collaborative Filtering, Content-based Filtering, Demographic Filtering

I. INTRODUCTION

Recommendation techniques are best known for their use on e-commerce websites, where they use input about a customer's interests to generate a list of recommended items [1]. Many applications on the internet studies the behavior of customers on items they have purchased, items viewed and user's past ratings which then finally recommend their users based on their interest. At e-commerce stores, they use recommendation algorithms to customize the online store for each customer in order to attract them by frequently displaying items which matches the customer interests. RS have become a useful tool which provides users with personalized recommendations on items such as clothes, books, movies, music and shoes.

This paper includes three of the most commonly used filtering techniques for RS. They are Collaborative Filtering (CF), Content-Based Filtering (CBF) and Demographic Filtering (DF). The use of RS implementation significantly increased with time in different ways and in diverse areas on the internet. Online social networks present new opportunities as to further improve the accuracy of RSs. In real life, people often resort to friends in their social networks for advice before purchasing a product or consuming a service. Findings in the fields of sociology and psychology indicate that

humans tend to associate and bond with similar others, also known as homophily [2].

The rest of the paper is organized as follows. Section II discusses the related work of each of the RS filtering techniques. Section III analyses the current status of these filtering techniques. We conclude the paper in section IV.

II. RELATED WORK

This section discusses with reference to the related work on the three basic RS filtering techniques used in social networks and e-commerce stores for suggesting useful recommendations to the users based on the use of networked databases and suggests goods to the customers by helping them to make the choice what to purchase.

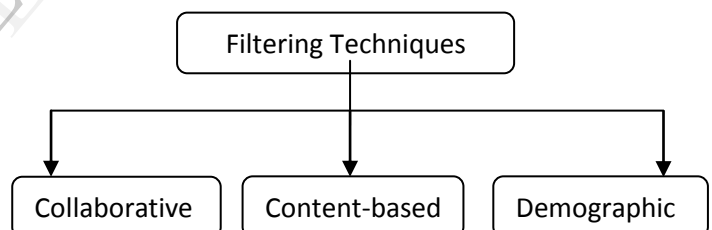


Fig. 1. Types of Filtering Techniques

Collaborative Filtering:

Collaborative filtering (CF) systems and computer-based recommendation are often related with the origin of the system called Tapestry. In Tapestry, with arbitrary text comments users were able to annotate documents and other users based on the comments of other users could then query. One of the main attribute of this system is that it allowed recommendations to be generated based on a combination of ideas of the input from many other users. Rather than filtering items based on content, making recommendations based on the opinions of like-minded users has become popularly known as collaborative filtering.

There have been many RS which use the CF method either in the business environment or in the academic world. This technique is still very useful even though it is the oldest recommendation technique and does not require to provide the representation of the object which can be easily read by the computers. The ability to make

serendipitous recommendations is the main advantage of CF. The problems with CF is that they only work best when the number of item rated per user is high and problems with new users and new products which have not yet been rated by the users but when there are many items to rate the problem which can occur is data sparseness that impedes finding users similar to the target one.

Content-based Filtering:

According to [3] Content-based filtering (CBF) is an outgrowth and continuation of information filtering research. The objects of interest are defined by their associated features in a CBF system. For instance, text recommendation systems like the newsgroup filtering system uses the words of their texts as features. Based on the features present in objects that the user has rated, a content-based recommender learns a profile of the user's interests which is called as "item-to-item correlation" and it derives the type of user profile depending on the learning method employed. Vector-based representations, neural nets and decision trees have all been used. Other system, in which the users rate the Web documents and assign them values from the binary "hot" and "cold" scale is [4]. Based on the ratings of the users, these ratings then serve as the basis for determining the probabilities of the words being in hot or cold documents. Then another system, the WebWatcher monitors the users' behaviors and choices of links on the WebPages in order to recommend links on the Web pages that the user will maybe visit in the future. In contrary to CF, the CBF is not so complex since only the analysis of the items that an independent user has bought or seen or must be done. It is not always possible to create the sufficient set of features, so in CBF each item is described by the features.

Demographic Filtering:

RS based on Demographic filtering (DF) classify users according to their demographic information and recommend services accordingly. In DF the user profiles are created by classifying users in stereotypical descriptions, representing the features of classes of users [5]. Demographic information identifies those users that like related services. Semi-trusted third parties use DF to recommend services by using data on individual users. DF creates categories of users which have similar demographic characteristics and then the cumulative buying behavior or preferences of users within these categories are being tracked. For a new user, recommendations are made by first finding which category he falls in and then the cumulative buying preferences of previous users is applied to that category which he belongs. Like collaborative techniques, demographic techniques also form "people-to-people" correlations but use dissimilar data. A collaborative and content-based technique requires a history of user ratings which is not of the kind required by Demographic approach.

TABLE I
RECOMMENDATION TECHNIQUES

Technique	Background	Input	Process
Collaborative	Ratings from U of items in I .	Ratings from u of items in I .	Identify users in U similar to u , and extrapolate from their ratings of i .
Content-based	Features of items in I	u 's ratings of items in I	Generate a classifier that fits u 's rating behavior and use it on i .
Demographic	Demographic information about U and their ratings of items in I .	Demographic information about u .	Identify users that are demographically similar to u , and extrapolate from their ratings of i .

III. CURRENT STATUS

This section provides the recent situation of the frequently used RS filtering techniques which are common, accepted and adopted at present.

1. **Collaborative Filtering (CF):** This filtering is probably the most widely implemented and most mature of the recommender systems. Collaborative systems are based on Collaborative filtering methods are based on collecting and analyzing a large amount of information on user's ratings, and generate new recommendations based on inter-user comparisons activities and predicting what users will like based on their similarity to other users technologies. A key benefit of the DF approach is not to rely on machine analyzable content and as a result it is able of precisely recommending complex items such as movies, songs, clothes, etc without requiring an understanding of the item itself. A lot of algorithms have been used in measuring items similarity in recommender systems. CF is based on the hypothesis that people who agrees in past will have the same opinion in future too and that they will like the related kinds if items they like in the past. These systems can be memory or model-based, comparing users alongside each other using relationship or other procedures in which a model is derived from the historical rating data and used to make predictions. CF approaches often suffer from three problems: cold start, scalability, and sparsity [6].
 - **Cold Start:** Very often these systems needs bulky amount of existing data on a user in order to make precise recommendations.
 - **Scalability:** In many of the environments that these systems make recommendations in, there are a large

number of users and products. Thus, a huge amount of computation power is often necessary to calculate recommendations.

- **Sparsity:** On major e-commerce sites, the quantity of items sold is really large. Of the overall database, the most active users will only have rated an undersized subset. Thus, even the most well-liked things have very a small number of ratings.

The biggest advantage of collaborative techniques is that they are entirely independent of any machine-readable representation of the things being recommended, and work well for complex things such as music and movies where variations in taste are dependable for much of the variation in preferences.

2. **Content-Based Filtering (CBF):** Another familiar approach when RS is content-based filtering. CBF methods are based on a profile of the user's preference and on a description of the item. In this filtering to describe the items, keywords are used and then a user profile is built to indicate the type of item this user likes. In short, these algorithms try to recommend items that are related to those that a user liked in the past. A variety of candidate items are compared with items earlier rated by the user and the best-matching items are suggested. To generate user profile, the system frequently focuses on two types of information: (i) a model of the user's preference (ii) a record of the user's interaction with the recommender system. Mostly, these methods use an item profile (i.e. a set of distinct attributes and features) characterizing the item inside the system. Based on a weighted vector of item features the system creates a content-based profile of users. The weights indicate the significance of each feature to the user and can be computed from independently rated content vectors using a variety of techniques.

Relative to collaborative filtering, content-based techniques also have the problem that they are restricted by the features that are explicitly related with the objects that they recommend. For example, content-based movie recommendation can only be based on written resources about a film: actors' names, plot summaries, etc. because the movie itself is unclear to the system. This puts these techniques at the mercy of the descriptive data existing. These systems rely merely on user ratings and can be used to recommend items exclusive of any descriptive data. Even in the presence of descriptive data, some experiments have found that collaborative recommender systems can be more accurate than content-based ones [7]. Direct feedback from a user, usually in the type of a like or dislike button, can be used to give different weights on the significance of definite attributes. A key concern with CF is whether the system is capable to study user preferences from user's actions about one content source and make use of them across other content types. So when the system is restricted to recommending content of the same type as the one the user is already using at present, the value from the recommendation system is significantly less than when other content types from other services can be

recommended. At the same time, various different presentations of the same wire-service story from different newspapers would not be useful.

3. **Demographic Filtering (DF):** It aims to classify the user based on personal attributes and make recommendations based on demographic classes. The users are divided into demographic classes in terms of their personal attributes. These classes serve as the input data to the recommendation process. The objective of this process is to find the classes of people who like a certain product. If people from class C like product s and there is person c (this user belongs to class C), who has not seen yet product s, then this product can be recommended to person c. The customers provide the personal data via surveys that they fill in during the registration process or can be extracted from the purchasing history of the users. This technique may not require collecting the complex data such as history of users' purchases and ratings. However, the weaknesses of DF are that the classification can be too general and this leads to lose the individuality of the users, this method uses data that are provided by users. This data can be either incomplete or untrue and the classification is created according to the customer's interest, which tend to vary over time. DF does not support the adoption of the user profile to changes.

TABLE 2
TRADEOFFS BETWEEN RECOMMENDATION TECHNIQUES

Technique	Pluses	Minuses
Collaborative Filtering (CF)	A. Can identify cross-genre niches. B. Domain knowledge not needed. C. Adaptive: quality improves over time. D. Implicit feedback sufficient	I. New user ramp-up problem. J. New item ramp-up problem. K. "Gray Sheep" problem. L. Quality dependent on large historical dataset. M. Stability vs. plasticity problem.
Content-based Filtering (CBF)	B, C, D	I, L, M
Demographic Filtering (DF)	A, B, C	I, K, L, M N. Must gather demographic information

IV. CONCLUSION

Recommender systems can create strategic advantages for companies that implement them and customer satisfaction can be achieved along with customer loyalty. Companies are more likely to be forced out of the market if it does not implement RS. In this paper we discuss three different types of RS but it is difficult to state that a particular RS is better than the other as quite simple systems can also be cheaper to implement although provide slower accuracy. However, virtually every RS will have problems when first implemented as there is inadequate data available on users and items. It is important to improve the accuracy when first implementing the RS because bad recommendations to users can reduce the effect RS have on sales.

Other potential problems in RS can be caused by malicious users and so the accuracy of the RS could be affected when large quantities of fake profiles are inserted. This might potentially lead to bad recommendations and thereby possibly contribute to a reduction of the effect RS have on sales.

REFERENCES

- [1] J.B. Schafer, J.A. Konstan, and J. Reidl, "E-Commerce Recommendation Applications," *Data Mining and Knowledge Discovery*, Kluwer Academic, 2001, pp. 115-153.
- [2] M. McPherson, L. Smith-Lovin, J.M. Cook, *Birds of a Feather: homophily in social networks*, *Annual Review of Sociology*, 2001, pp 415-444.
- [3] Belkin, N. J. and Croft, W. B, 'Information Filtering and Information Retrieval: Two Sides of the Same Coin?' *Communications of the ACM*, 1992, 29-38.
- [4] Pazzani, M., Muramatsu, J., Billsus, D. 'Syskill & Webert: Identifying Interesting Web Sites', *Proceeding of the 13th National Conference on Artificial Intelligence*, 1996, 54 – 61
- [5] Montaner, M., Lopez, B. and De la Rosa J.L. 'A Taxonomy of Recommender Agents on the Internet', *Artificial Intelligence Review*, Kluwer Academic Publisher, 2003, 285 – 330.
- [6] Breese, J. S., Heckerman, D. and Kadie, C, 'Empirical analysis of predictive algorithms for collaborative filtering'. *Computer Networks and ISDN Systems*, 1998, pp. 43-52.
- [7] Alspector, J., Koicz, A., and Karunanithi, N, 'Feature-based and Clique-based User Models for Movie Selection: A Comparative Study'. *User Modeling and User-Adapted Interaction* 7, 1997, pp. 279-304.

IJERT