International Journal of Innovation Engineering and Science Research
(ISSN: 2581-4591)
www.ijiesr.com

Categorization of recommender system methods

Chiheb Eddine Ben Ncir
Management Information System
College of Business, University of Jeddah
Jeddah, Saudi Arabia

ABSTRACT

Recommender systems has become a necessary requirement in several application domains due to information overload. The development of web applications and social networks has made data widely and largely available. Filtering necessary and important data has become a non-trivial task. Recommender systems can help users to filter relevant data and provide them with the best suggestions. To develop such automatic system, several methods have been used to filter data and to build suggestions. Usually, these methods are categorized into three main categories which are: content-based, collaborative-based and hybrid methods. In this paper, we give a new categorization of recommender system methods where methods are categorized into six categories. We add three new categories which are multi-criteria recommendation, context awareness-based recommendation and social network-based recommendation.

Keywords*—recommender systems; collaborative filtering; content-based filtering; Hybrid methods*

I. Introduction

The web has become the largest accessible information space in the world [1]. In this space, information in its various forms grows exponentially which results in large amount of data available for users. It becomes more and more difficult for regular Internet-users to successfully find items related to their tastes and preferences. For example, an internaut browsing an online filmor movie does not wish to go through thousand of uninteresting movies before finding a preferred one. The increasing number of choices makes it harder for users to find exactly the preferred item. Fueled by this need to combat information overload, recent years have witnessed an increase of research in the area of recommender systems.

Recommender Systems are software tools providing users with suggestions related to their tastes and preferences. They help users to deal with information overload by giving suggested items to a user related to his tastes and expectations. An item can be an article to read, a product to buy, a piece of music to listen to, a movie to watch or a web page to consult. The objective is both to minimize his time spent on research, but also to suggest relevant items that he would not have spontaneously consulted and thus increase his overall satisfaction.

Recommendation systems have become an important research area with the publication of the first articles in the collaborative filtering field. The academic literature has introduced the term collaborative filtering by “Tapestry” [2] system that represents one of the first recommendation systems. It was developed in 1992 by the research center of “Xerox” in the United States. A few years later, with the growth of the Internet, and web applications, there have been a keen interest in recommendation systems that have developed in different areas of application, such as movies recommendation like Netflix or MovieLens [3], Musical titles Recommendation like Last FM, news recommendation like Yahoo! News, products recommendation introduced on e-commerce sites like Amazon [4], bibliographic citations recommendations systems like Tech Lens [5], courses recommendation systems introduced in e-learning sites, friends recommendation systems like Twitter or Facebook.

These applications show that recommender system sare receiving an important intention [6] by the research community. Indeed, there has been dedicated conferences and workshops related to this field such as the ACM Recommender Systems (Rec Sys) established in 2007. In addition, sessions dedicated to recommender systems are frequently included in the more traditional conferences in the area of data bases, information systems and adaptive systems. Recent interesting books [7] related to recommender systems has been recently published introducing recommender system’s techniques. Moreover, there have been several special issues in academic journals covering research and developments in the recommender system field.

This paper gives an overview of recommender systems methods and techniques which are adopted in several existing systems. We give a classification of these methods based on the approach used to give recommendations. The rest of the paper is organized as in the following: Section II presents main components of recommender systems, then Section III presents main concepts and approaches of recommendation systems while giving a classification of existing methods. Finally Section IV gives conclusions and future works.

II. RECOMMENDER SYSTEMS COMPONENTS

Recommender systems are information processing systems that actively gather various kinds of data in order to build their recommendations [7]. Data is primarily about the items to suggest and the users who will receive these recommendations. In general, these data are dependent to the used recommendation technique. However, data used by recommender systems can refer to three kinds of objects [7]: items, users, and Preferences.

- **Users** : People who accesses the system , entering his demographic information, his interests and other personal information.
- **Items**: Products which the system can recommend are referred to as "items". These items correspond to the need of the user, including any product likely to be seen (movies in online TV sites such as Netflix), sold (book in the e-commerce sites such as Amazon.com), listened (music) or read (such as scientific article in a digital libraries).
- **Preferences**: In order to provide personalized recommendations adapted to the needs of the active users, recommender systems must collect personal preference information. Two categories of preferences data exist, the first is the explicit data [8] expressed by users during their navigation activities. For example, providing a note on a scale of predefined values (notes that users indicate on products they buy on the Internet for example), express an opinion on an object (example of the button "Like" on Facebook), etc. The second category is the implicit data [9] collected by observing user behavior and activity. The activity may be web pages viewed, the user's history of purchases, click-stream data, etc.

III. CATEGORIZATION OF RECOMMENDER SYSTEM METHODS

To get an earlier overview of the different types of recommendation systems, a taxonomy provided by authors in [10 has been considered a benchmark for classifying these systems. Three main methods are defined:

- Content-based recommendation system: The relevance of the recommended items is estimated by the similarity between the features or content of the items, and the profile of the user reflecting its needs in terms of content.
- Collaborative-based recommendation system: The system recommends items seen by other users with similar tastes or common interests.
- Hybrid recommendation system: The system exploits the complementary of the two previous approaches by combining them.

In addition to these recommendation categories, we added new recommendation categories which are multi-criteria recommendation, context awareness-based recommendation and social network-based recommendation. These recommendation system methods are presented in more detail in the following subsections.

A. *Content-based recommendation methods*

Content-based method tries to recommend items that are similar to items previously liked by a specific user as described in Figure1. To return recommendations, the system needs the following information:

- **item profile**: the item profile is a set of features that represent the content of the recommended item. For example, in a research article recommendation system, the attributes adopted to describe an article are title, author, year, venue, …, etc. Content-based method depends strongly on the nature of the data to be represented and it finds difficulties for the representation of the items, which decreases its accuracy from one type of resource to another. Textual data is the best represented and with which the method gives the best accuracy, whereas other data types, i.e. non-textual data such as videos and images, give poor precision due to the difficulty in extracting characteristics. Generally, for non-textual data, the approach uses metadata to represent its content.

- **user profile**: The user profile build a description of user needs and preferences. From the user profiles and descriptions of the items to be recommended, this content-based recommender system analyzes their similarity, then it builds a list of suggestions to present to the user.

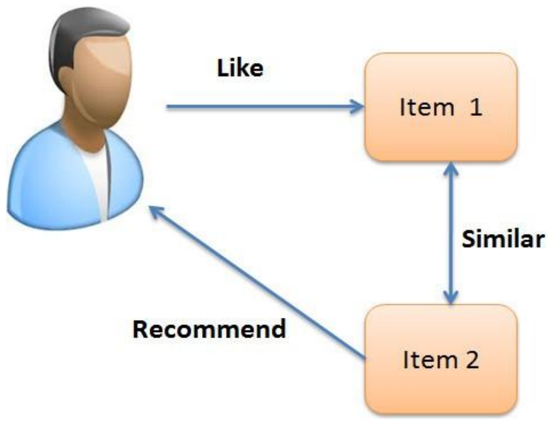

Fig1.Content-based recommendation: recommend items that are similar to items previously preferred by a similar user

One major drawback of the content-based filtering (recommendation) is the new user cold start problem. When a new user requests a recommendation, there is only little information about this user. Therefore, it will be very difficult for the recommender system to produce relevant recommendations. Moreover, content-based filtering usually suffers from the over-specialization problem. The system will infinitely suggest items that are highly matching with the user profile. It will not find unrelated items to the one already appreciated.

B. Collaborative filtering-based recommendation methods

Collaborative filtering is one of the most explored methods in the area of recommendation. The main advantage of this approach is that it doesn't require any description of the items to recommend. Therefore, this method allows to recommend complex objects without having to analyze them, thus most popular online music, and movie recommendation services such as Netflix and Last fm apply it.

The main idea of these methods is to recommend items based on the opinions and the ratings of other similar users. The opinions of users can be obtained explicitly from the existing users or by using some implicit measures. Explicit opinions are directly provided by the user. For example, when a new user logs in, many systems begin by asking the user to note multiple items so that they can subsequently recommend items that match them; this is the case of Amazon recommendation system. It is then possible to note the items. These notes can be:

- binary note : the decision of the user is to choose whether the object is "good" or "bad".
- digital note : generally ranging from 1 to 5; 5 reflects the fact that the object is very pleasing to the user. These notes are often represented by a number of stars as on Amazon system as described in Figure 2.
- ordinal note: the user must choose from a list of terms the one that is the most adapted for his feeling towards the item in question. An example of a list can be "Very good, good, average, bad, very bad".
- descriptive note: the user chooses one or more terms that describe his opinion on the object. For a movie, these terms can be "exciting", "uninteresting", etc.

Implicit opinion gathers all the information that can be extracted from the usage's patterns. For example, the number of visits to a page, the number of views on a video, the time spent on a page (Dwell time) which can reflect the fact that a user has taken the time to read or not the content of this page. In a collaborative filtering system, as described in Figure 3, the data are represented as "User x Item" matrix as described in Figure 3, where the rows represent users (Joe, Dalia, Eric, Helena) and columns correspond to the items (the set of book list: Harry Potter, Springer, Rever.). Users provide their opinions on the items in the form of notes.

According to [11], there are two sub-categories of collaborative filtering: memory-based and model-based algorithms.

Fig2. Explicit user opinion's in amazon recommendation system

The memory-based methods use the entire evaluation matrix and provide recommendations based on the relation- ship between the active user and item and the rest of the rating matrix. On the other hand, model-based methods use the evaluation matrix to fit a parameterized model, which will then be used for predictions. We give in the following basic concepts of each sub-category.

For the first category, Memory-based methods, it can be divided into two kinds: user-based CF and item-based CF [12]. User-based CF explores the relationship between rows(users) in the rating matrix. It firstly calculates the similarities between the active user and the other users. The most popular measures are the Pearson correlation [3] and the cosine-based measure [13] which are described by Equation (1) and Equation (2) where $Cos(u_a, u_b)$ and $CorrP(u_a, u_b)$designate the similarities calculated respectively with the Pearson correlation coefficient and the cosine-based measure, between two users u_a and u_b, I_a and I_b respectively represent all items rated by $u_a$and u_b, $v(u_a)$ represents the average of notes of u_a, $v(u_a, i)$ represents the note of $u_a$on item i and I_c represents the co-noted items between the active user $u_a$and the user u_b.Users with high similarities are selected as neighbors of the active user. Then the system utilizes the neighbors' ratings on a specific item to calculate the weighted average [3] which is regarded as the predicted rating, treating the respective similarities as weights. This method is described in Equation 3, where $Pred(u_a, i_k)$ is the prediction note of the active user $u_a$on the item i_k, $U_a$is the set of nearest neighbors having already noted the item i_k, $Sim(u_a, u_b)$ means the similarity value between $u_a$and $u_b(u_b \in U_a)$ that can be the cosine similarity ($Cos(u_a, u_b)$ or the Pearson correlation similarity ($CorrP(u_a, u_b)$). The recommender system ranks all the items according to their predicted ratings and selects Top-N items as recommendations.

				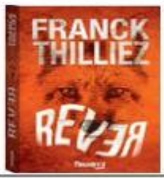
Joe	1	?	3	?
Dalia	1	?	4	3
Eric	1	5	?	5
Helena	5	?	1	1

Fig3. Collaborative filtering matrix: rows represent users and columns represent items

The method based on the memory is concentrated on the items [14]. This method analyzes the matrix "User x Item" to identify the relationships between the items and use these relationships to calculate predictions. The assumption is that the user would be interested in the items similar to items he previously appreciated (similar in terms of ratings assigned by the user). Hence the similarities among items are firstly calculated using the same similarity measures with the user-based CF. After choosing the neighbors for the target item and calculating the weighted average, the predicted rating on this item is obtained.

$$Cos(u_a, u_b) = \frac{\sum_{i \in I_c} v(u_a, i) * v(u_b, i)}{\sqrt{\sum_{i' \in I_a} v(u_a, i')^2} * \sqrt{\sum_{i' \in I_b} v(u_b, i')^2}} \quad (1)$$

$$CorrP(u_a, u_b) = \frac{\sum_{i \in I_c} (v(u_a, i) - \overline{v(u_a)})(v(u_b, i) - \overline{v(u_b)})}{\sqrt{\sum_{i \in I_c} (v(u_a, i) - \overline{v(u_a)})^2} \sqrt{\sum_{i \in I_c} (v(u_b, i) - \overline{v(u_b)})^2}} \quad (2)$$

$$Pred(u_a, i_k) = \overline{v(u_a)} + \frac{\sum_{u_b \in U_a} Sim(u_a, u_b) * (v(u_b, i_k) - \overline{v(u_b)})}{\sum_{u_b \in U_a} Sim(u_a, u_b)} \quad (3)$$

The advantage of the memory-based methods is the simplicity of the implementation and the integration of new data into the system. However, this approach suffers from the cold start issue. It requires rating data in order provide recommendations. When a new item is added, there is no enough information about that item. Consequently, this item cannot be recommended to anyone before it gets enough ratings. For example, in the online movies, when a new movie is uploaded, there is no rating that is given to that movie. So, this new movie cannot be recommended by the system until collecting enough ratings and evaluations by a group of users. Therefore, this category of methods cannot provide recommendations before a target user has given enough ratings for several items. Data sparsity is also one of the main problem of memory-based methods when the evaluation space is dispersed, it is difficult to identify reliable neighbors (from the co-scored items) and consequently the system performance decreases. For example, considering movie domain, the evaluation matrix includes thousands of movies and millions of users. However, a given user evaluates only a small number of movies. As a result, the evaluation matrix has many empty cells, which makes it difficult to calculate the similarity between users. In addition, them emory-based methods do not allow scaling. Indeed, when the number of users and items in the system becomes large, the generation of recommendations requires a very high processing time.

On the other side, the second subcategory is model-based methods. A variety of model-based recommendation methods were developed to address the scalability and the real time performance problems [15] of user-based methods by using dimensionality reduction techniques or clustering in order to remove users or non-representative items. Therefore, the user-item representation space is smaller, and the missing data rate are less important compared to the original representation space. Model Based algorithms received an important attention by the recommender system community.

C. Hybrid recommandation methods

Each of the presented recommendation approaches has strengths as well as weaknesses [16]. Therefore, hybrid system [17] combines two or more recommendations techniques to address the disadvantages of each technique and take advantage of their strengths. There are various possible combinations which is classified [18] into seven categories:

- Weighted: hybrid recommender combines the results of all available recommendation techniques to compute a score for a recommended item.
- Switching: the system switches between recommendation techniques depending on the current situation.
- Mixed: recommendations from several techniques are presented simultaneously to the user.
- Feature combination: features from different recommendation data sources are thrown together into a single recommendation algorithm.
- Cascade: one recommender refines the recommendations given by another.
- Feature augmentation: the output from one technique is used as an input feature to another.
- Meta-level: the model learned by one recommender is used as input to another.

D. Multi-Criteria recommendation method

Single-rating collaborative filtering recommender systems have shown their effectiveness to provide acceptable automatic recommendations. However, they do not include different attributes and dimensions of an item in the process of recommendation. Multi-criteria recommender systems have been proposed to address this issue. Several works show that using multi-criteria ratings, instead of a single rating, can largely improve the performance of the recommender system and may offer a more relevant and reliable recommendation. An example ofa multicriteria rating system [19] is Zagat's system that provides three criteria for restaurant ratings (e.g., food, decor and service),"Buy.com" that provides multi-criteria ratings for consumer electronics (e.g., display size, performance, battery life, and cost), and "Yahoo" Movies that show each user's ratings for four criteria (e.g., story, action, direction, and visuals).

E. Context awareness-based recommendation method

One of the most cited definitions of context is the definition of [20] that defines context as "any information that can be used to characterize the situation of an entity. An entity could be a person, a place, or an object that is considered relevant to the interaction between a user and an application, including the user and the application themselves." The context information such as time or location has been recently considered in existing recommender systems. The contextual information provides additional information for recommendation making, especially for some applications in which it is not sufficient to consider only users and items, such as recommending a vacation package. It is also important to incorporate the contextual information in the recommendation process to be able to recommend items to users in specific circumstances. For example, using the temporal context, a travel recommender system might make a very different vacation recommendation in winter compared to summer. Nowadays, when the mobile devices are getting popular and taking part in our lives, this kind of recommender systems are especially urgent. Using GPS, 3G access to internet and other technologies recommender systems can rapidly get any information about location of a user and the user himself.

F. Social network-based recommendation method

Social network analysis has been used in recommender systems as a result of the important growth of social networking tools in Web-based systems in recent years. Data is automatically collected from social networks is a real source of information for a recommendation system. Such system is based on the presence of a community of users linked by social links. In social platforms, these recommendation systems make it possible to recommend a whole range of information. Examples include users to follow, specific publications, multimedia elements, groups (sub-communities)to integrate.

IV. Conclusion

In this paper we presented the main recommendation methods, namely: content-based methods, memory-based and model-based CF, hybrid methods, multi-criteria recommendation methods, context awareness-based recommendation methods and social network-based recommendation methods. There is no method category that is superior to all others or any method that can be successfully applied for any application of recommender systems. Each technique has some advantages but also some disadvantages. The choice of one technique is based on the problematic, the application context and the availability of data. An interesting improvement of this work is to prepare a guide in which we can specify key factors affecting the choice of a recommendation method given a list of different target applications.

References

[1] Akshay Kumar Chaturvedi, Filipa Peleja, and Ana Freire. "Recommender System for News Articles using Supervised Learning". In: arXiv preprint arXiv:1707.00506.2017.

[2] David Goldberg et al. "Using collaborative filtering to weave an informa- tion tapestry". In: Communications of the ACM 35.12.1992, pp. 61–70.

[3] Jonathan L Herlocker et al. "An algorithmic framework for performing col- laborative filtering". In: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. ACM. 1999, pp. 230–237.

[4] Greg Linden, Brent Smith, and Jeremy York. "Amazon. com recommenda- tions: Item-to-item collaborative filtering". In: IEEE Internet computing 7.1.2003, pp. 76–80.

[5] Sean M McNee et al. "On the recommending of citations for research pa- pers". In: Proceedings of the 2002 ACM conference on Computer supported coop- erative work. ACM. 2002, pp. 116–125.

[6] Francesco Ricci, Lior Rokach, and Bracha Shapira. "Introduction to recom- mender systems handbook". In: Recommender systems handbook. Springer, 2011, pp. 1–35.

[7] Dietmar Jannach et al. Recommender systems: an introduction. Cambridge University Press, 2010.

[8] J Ben Schafer et al. "Collaborative filtering recommender systems". In: The adaptive web. Springer, 2007, pp. 291–324.

[9] Bamshad Mobasher, Xin Jin, and Yanzan Zhou. "Semantically enhanced collaborative filtering on the web". In: Web Mining: From Web to Semantic Web. Springer, 2004, pp. 57–76.

[10] Gediminas Adomavicius and Alexander Tuzhilin. "Toward the next gener- ation of recommender systems: A survey of the state-of-the-art and possi- ble extensions". In: IEEE transactions on knowledge and data engineering 17.62005, pp. 734–749.

[11] John S Breese, David Heckerman, and Carl Kadie. "Empirical analysis of predictive algorithms for collaborative filtering". In: Proceedings of the Four- teenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc. 1998, pp. 43–52.

[12] Kwanghee Hong, Hocheol Jeon, and Changho Jeon. "Personalized research paper recommendation system using keyword extraction based on user- profile". In: Journal of Convergence Information Technology 8.16.2013, p. 106.

[13] Badrul Sarwar et al. Application of dimensionality reduction in recommender system-a case study. Tech. rep. DTIC Document, 2000.

[14] Badrul Sarwar et al. "Item-based collaborative filtering recommendation algorithms". In: Proceedings of the 10th international conference on World Wide Web. ACM. 2001, pp. 285–295.

[15] Mukund Deshpande and George Karypis. "Item-based top-n recommendation algorithms". In: ACM Transactions on Information Systems (TOIS) 22.1 2004, pp. 143–177.

[16] Toine Bogers and Antal Van den Bosch. "Recommending scientific articles using citeulike". In: Proceedings of the 2008 ACM conference on Recommender systems. ACM. 2008, pp. 287–290.

[17] Robin Burke. "Hybrid recommender systems: Survey and experiments". In: User modeling and user-adapted interaction 12.4 2002, pp. 331–370.

[18] Robin Burke. "The adaptive web. chapter Hybrid web recommender sys- tems". In In: Brusilovsky P., Kobsa A., Nejdl W. (eds) The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Springer, Berlin, Heidelberg 2007.

[19] Gediminas Adomavicius and YoungOk Kwon. "Multi-criteria recommender systems". In: Recommender Systems Handbook. Springer, 2015, pp. 847–880.

[20] Anind K Dey, Gregory D Abowd, and Daniel Salber. "A conceptual frame- work and a toolkit for supporting the rapid prototyping of context-aware applications". In: Human-computer interaction 16.2 2001, pp. 97–166.