Recommendation Systems: a review

[1]Amir Hossein Nabizadeh Rafsanjani, [2]Naomie Salim, [3]Atae Rezaei Aghdam, [4]Karamollah Bagheri Fard
[1, 2, 3, 4] *Faculty of Computing, Universiti Teknologi Malaysia (UTM), Skudai, 81310 Johor, Malaysia*

ABSTRACT
This article presents an overview of recommendation systems and illustrates the present generation of recommendation techniques that are usually categorized into the following three main classes: Collaborative Filtering (CF), Content-Based Filtering (CBF), and Hybrid Recommendation approaches. This paper also describes several weaknesses of current recommendation techniques.

Keywords: *Recommender systems, Collaborative Filtering, Content-Based Filtering, Hybrid approach*

I. INTRODUCTION

Recommendation is becoming one of the most important methods to provide documents, merchandises, and cooperators to response user requirements in providing information, trade, and services that are for society (community services), whether through mobile or on the web [1].

The quantity of data and information has been increasing daily which causes overloading of information and data. At this time, finding the customers' requirements and tendencies became important as this problem changed into the big problem. One of the innovations which helped people a lot were the engines for search (search engines) and they were somewhat as a solution for this problem.

Anyway, the information could not be personalized by these engines. System developers introduced a solution for this problem that named recommendation system.

This system use to sort and filter information, data, and objects. Recommendation systems utilize users' idea of a society or community to assist for realizing effectively users' tendency and also demands in a society from a possibly onerous set of selections [2].

The main aim of recommendation system is creating significant suggestions and recommendations information, products or objects for users' society that users could interest them. For instance, book recommendation on Amazon site, Netflix that recommend movies that use recommendation systems to identify users' tendencies and subsequently, attract users more and more [3].

There are a lot of different methods and algorithms which can assist recommendation systems to create recommendations that are personalized. All of the recommendation approaches can be divided in these three categories which are very famous:

- Content-based recommending: This method suggests and recommends objects and information which are comparable in content to objects that the users have interested previously, or compared and matched to the users' characteristics.
- Collaborative Filtering (CF): Collaborative Filtering systems suggested and recommended objects and information to a user according to the history valuation of all users communally.
- Hybrid methods: Hybrid methods are a combination of Content-based recommending and Collaborative Filtering (CF) methods [3].

This paper is organized as follows: Section 2 includes a review of collaborative filtering recommender systems literature to highlight the differences among them and indicates the advantages and disadvantages of collaborative filtering. Next, Section 3 illustrates the content base filtering (CBF) and also the pros and cons of utilizing this algorithm. Section 4 explains the different types of hybrid recommendation systemFinally, the conclusion of this study is explained during Section5.

II. COLLABORATIVE FILTERING (CF)

Collaborative filtering (CF) is one of the most famous methods for categorization the objects and has proved which CF is very effective for forecasting customer precedence in choice of objects. This method or Collaborative filtering (CF) is flourished in the middle of 1990s with scattering of services which utilized recommendation systems and presented online, like Netflix, Amazon, Elsevier. Collaborative filtering (CF) is designed to work on enormous database [4].

Collaborative filtering (CF) attempt to mechanize "word-of-mouth" recommendation procedure that means the objects suggested to customer according to how customers that have similar interests, categorized these objects [7].

At first, Goldberg et al used Collaborative filtering (CF) for introducing their filtering system that gives ability to customer for explanation their e-mails and documents [5]. Other customer can ask for documents that elucidated by specific people, but recognition of these people was left to customers. Collaborative filtering (CF) methods mechanize this procedure of identification close neighbors of customer that is active.

Collaborative filtering (CF) algorithms utilize patterns which demonstrate customers' precedence and interaction for accordance them to customers share information and documents which are similar. After recognition a match that is possible, the suggestions and recommendations are generated by algorithm. The patterns which Collaborative filtering (CF) algorithms utilize for precedence can extract from customer, directly.

It is an example of Collaborative filtering (CF) that Amazon Website use where users and customers are required to sort an object from A to E. After collection implicitly or explicitly customers' opinion, the Collaborative filtering (CF) usually use matrix for rating customers' object. As it is shown in next figure, the most number of cells are vacant, Because of it is impossible for a customer to chose, buy or categorize all of the objects that are exist in a system. Collaborative filtering (CF) algorithms utilize to anticipate values for vacant cell in matrix.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
User 1	A	B	E		C		E		A	
User 2	A		B	C		E		B		D
User 3		A		D	E		C		E	B
User 4	D		C	A	E			B	D	C
User 5	B	C		D		E	A			
User 6	B	E		D	A			C	D	E

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
User 1	1	1	1		1		1		1	
User 2	1		1	1		1		1		1
User 3		1		1	1		1		1	1
User 4	1		1	1	1			1	1	1
User 5	1	1		1		1	1			
User 6	1	1		1	1			1	1	1

Figure 1: Two instances of users' object matrix which includes ten objects and six users.

Collaborative filtering (CF) algorithms usually separated into two parts:

- Model-based algorithm.
- Memory-based algorithm.

1.1. Memory-based Collaborative Filtering

Another name of the algorithms of Memory-based is lazy recommendation algorithms. They postpone the calculative attempts for forecasting a customers' precedence for an object to the time that customers ask for a collection of recommendations.

The training stage of algorithm of memory-based includes storing the entire customers' ranking into memory.

There are two different memory-based recommendations that are according to k-Nearest neighbor algorithm [8]:

- Item/Object-based filtering.
- User/Customer-based filtering.

Item/Object-based filtering recommended by Sarwar et al at 2001 [9]. It mostly focuses on understanding the most similar items/objects. Items/objects are regarded for similarity when the same collection of customers has ranked or bought them highly. For every object that belongs to the customer who is active, the neighborhood of most likely objects is recognized. Every top k-neighbor is put on a list of applicants together with its likeness to the object of user who is active. The scores of similarity of objects that happening several times in applicant list are sum. The applicant list is categorized on these accumulated likenesses scores and the top N suggestions and recommendations are presented to customer [9, 10].

User/Customer-based filtering match the customer/user who is active versus the ranking matrix for finding the neighbors of active customer with which user that is active have a past concurring. At first, all of the neighbors identify, the entire object in profile that are belonging to the neighbors which are strange for customer that is active are regarded as suggestion and recommendation that are possible and classified in the neighborhood by their frequency. A rate according to their value accumulate of these frequencies generate recommendations [11].

1.2. Model-based Collaborative Filtering

Another name of this model is Eager recommendation algorithms, Model-based Collaborative Filtering algorithms do majority of work that is hard in the training stage, where these algorithms build a forecasting model of problem in recommendations. Producing the suggestions or recommendations is uncomplicated and fast issue of applying the derived model [12].

Model-based Collaborative Filtering has two probabilistic model:

- Cluster Model
- Bayesian Network Model

1.2.1. Cluster Model

One of the models that is accepted for Collaborative Filtering (CF) is Cluster model. Another name of this model is multinomial mixture model where the probability is provisionally independent from membership votes in a class C variable accepting several comparatively small numbers of values that are discrete. The Cluster model idea is that there are several definite groups or kind of users that taking a collection of precedence and preferences that are common among them. In the certain classes, the precedence that are related to the different items are independent. This model explains the probability of joint probability of votes and class to a collection of marginal and conditional distribution [12].

$$\mathbf{Pr}\ (\mathbf{C}=\mathrm{C1V1,C2V2,\ldots,Vn})= \mathbf{Pr}\ (\mathbf{C}=\mathrm{C}) \qquad \mathbf{Pr}\ (\mathbf{Vi}\ |\mathbf{C} = \mathrm{C})$$

The left side of this formula is observing probability of a person of specific class and a collection of votes. This model has several parameters include **Pr (C**=C**)** that is class membership probability and **Pr (Vi |C =** C**)** condition probability of votes that are estimated from a collection of users' vote training. We cannot see the variables that are located in user database then we have to use methods that can learn parameters with hidden variables for modeling.

1.2.2. Bayesian Network Model

Another model can be use for Collaborative Filtering (CF) is Bayesian Network Model with a node similar to every item that is located in domain. The states of a node are based on possible value for every item. We also have one state based on "NO VOTE" where there is not a logical interpretation for data that are lost in domains.

After that, we use an algorithm to train data for learning Bayesian Network Model and where data are lost indicate by "NO VOTE". The algorithm that is used for learning, search over different structures of model based on dependencies for every item [12].

Therefore, every item has a collection of parent items which can forecast votes very well. Every table that is related to condition probability is presented by decision tree which is coded the probabilities that are conditional for that node. The learning algorithms is more discussed in Chickering et al. at 1997 [13].

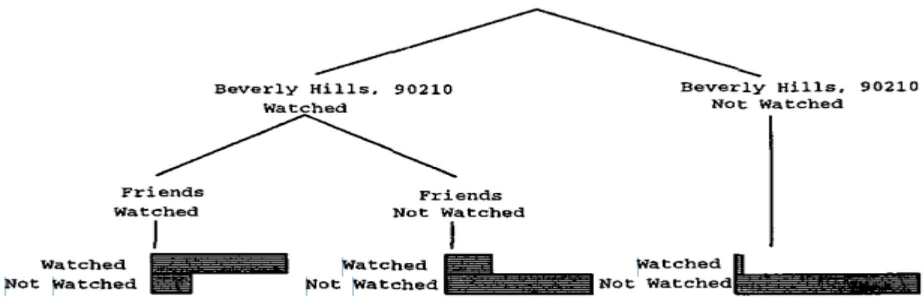

Figure 2: A decision tree for whether an individual watched "Melrose Place", with parents "Friend's", and "Beverly Hills, 90201". The bar charts at the bottom of the tree indicate the probabilities of watched and not watched for "Melrose Place", conditioned on viewing the parent programs [12].

1.3. Pros and Cons of Collaborative Filtering

Collaborative Filtering (CF) algorithms has several pros, like capability for taking an object/item quality or defect into an account when suggesting objects/items, particularly in explicit customer rankings. For example, a local music band could fall into the same genre of music a rock band that is famous in all over the world, but this item does not assurance which they have same level of quality. This subject demonstrates that objects/items identification quality is obvious pros of Collaborative Filtering (CF). Collaborative Filtering (CF) can hinder deficient suggestions and recommendation by taking the precedence of customers which are actual into an account. Second pros is which the Collaborative Filtering (CF) algorithms are particularly applicable and useful in domains where the analysis of content is very expensive or difficult, like music and film suggestion, without demanding any domain of knowledge [14].

Although the Collaborative Filtering (CF) algorithms has several pros and the quality level of Collaborative Filtering (CF) algorithms improve during the time, but the most important problem is the phase of startup in recommendation system, as there are many objects and items are provided in the system while there are few customers and few or no rankings. This problem named “cold start” and means that recommendation system cannot produce any suggestion or recommendations [15]. Remedies for solving this problem involve seeding the system by utilization other data sets, and using algorithms of recommendation system that are different in startup phase which do not suffer from “cold start” problem. Even after obtaining more ranking from customers, scantiness of the customer-object matrix can still be a problem for Collaborative Filtering (CF).

Second problem named “gray sheep” with regarding to Claypool et al. at 1999, that is a description about the hardship of recommendation system for people who are not belong to the part of an obvious group [16]. Collaborative Filtering (CF) is useful and work very well for customer and user who are fit into a particular group with a lot of neighbors that are similar [17].

Scalability is the next challenge of CF. When the number of objects and customer increase, the traditional form of Collaborative Filtering (CF) suffers critical from scalability problem. For instance, with a enormous population of customers and also big umber of objects and items, then the intricacy of Collaborative Filtering (CF) will increased. At this time, we need many systems to response urgently for online demands that we require a higher level of scalability of a Collaborative Filtering (CF).

Another challenge that Collaborative Filtering (CF) is faced is synonymy. This problem related to inclination of numerous of very similar objects to have distinctive names. Recommendation systems usually are not capable to find this problem then faced with these objects differently. For instance, “adult automobile” and “adult car” are different statement but both of them allude to the similar object. In fact, the performance of Collaborative Filtering (CF) will decrease by propagation of synonyms.

Shilling Attacks can be another challenge for recommendation systems. It means when every item or object can be ranked by every customer, in comparison with other objects that belonging to other people, customers maybe give higher rank to own objects and items or even give negative rate to competitors’ products. That’s why in many cases, Collaborative Filtering (CF) systems must establish safety measure to dissuade customers and users from Shilling attacks.

III. CONTENT-BASED FILTERING

The second famous recommendation algorithm is Content-based recommendation algorithms. Another name of these algorithms is content-base filtering.

These algorithms can be seen as an extended work that is performed on filtering of information [6]. Normally, the methods of content-based filtering regard to build several type of representation of content in system and then learning customers’ precedence profile. Then, the representations of the content are matched opposite customers’ precedence profile to discover the objects which are most related to that customer. As with Collaborative Filtering (CF), the representations of customers’ precedence profile are models which are long-term, and also we can update precedence profile and this work become more available [14].

Generally, the method of recommendation by content-based filtering has problem, where representation of document must be matched to representation of customer on similarity of text or problem of machine learning when the content which is a text of representations are unified that are utilized to train an algorithm of forecasting.

There are several instances of the machine learning in Mooney and Roy at 2000 and also Lang at 1995 [18, 19].

1.4. Advantages and Disadvantage of Content-based Filtering

One of the most obvious advantages of content-based filtering algorithms is these algorithms don not need to domain of knowledge. It is adequate to gather feedback from customers about their precedence.

Next advantage of content-based filtering algorithms that we can consider to it is, these algorithms are better than Collaborative Filtering (CF) at finding locally similar objects. Because the explicit focus of content-based filtering algorithms is on similarity of text. However, this item can be a defect in domains where analysis of content in large number is impractical, impossible or difficult, like music and movies. The tendency of algorithms of content-based filtering is get stuck in a “well of similarity” [20], where they suggest objects only from a restrict theme scope. Then the recommendations that are serendipitous can be very difficult to achieve.

IV. HYBRID RECOMMENDATION SYSTEMS

Hybrid recommendation systems are adjusted for joining Content-based and Collaborative Filtering (CF) that control by one framework, and increase the benefits and also decrease the weaknesses of both techniques. Therefore, hybrid recommendation systems work on characteristics that are related to both. Indeed, there are many approaches that we can unite Content-based to Collaborative Filtering (CF). Several methods for combining Content-based to Collaborative Filtering (CF) list by Bruke at 2002 [14]. Pazzani introduced methods that discovered similitude among customers by building a content–based profile [21]. For example, Fab which recommends Web pages to its customers and users is one of hybrid recommendation systems [22]. The various hybrid recommendation systems are suggested for citation of research articles by McNee et al. at 2002 and alsoTorres et al. at 2004 [23, 24].

Several of the combination approaches that are used for building hybrid recommendation systems are as follows:

- **Mixed**: this method point to the suggestions and recommendations which are recommended from a set of various recommendation systems, are presented simultaneously.
- **Weighted**: Production a single recommendation by utilization of the votes and rates that are produce by some recommendation approaches.
- **Feature combination**: The characteristics which are relate to various recommendation data resources are get together into a single recommendation system algorithm.
- **Cascade**: One of the recommendation systems purify the suggestions and recommendations that are presented by another recommendation system.
- **Feature augmentation**: the results from one approach are utilized as input data and characteristics for another recommendation method.
- **Meta level**: the approach that is learned by one recommendation system is utilized as a input for another approach.
- **Switching**: in this method, recommendation system switches among recommendation approaches according to the current situation [14].

For example, the PTV system utilizes mixed technique to assemble a recommendation program of television viewing [25].

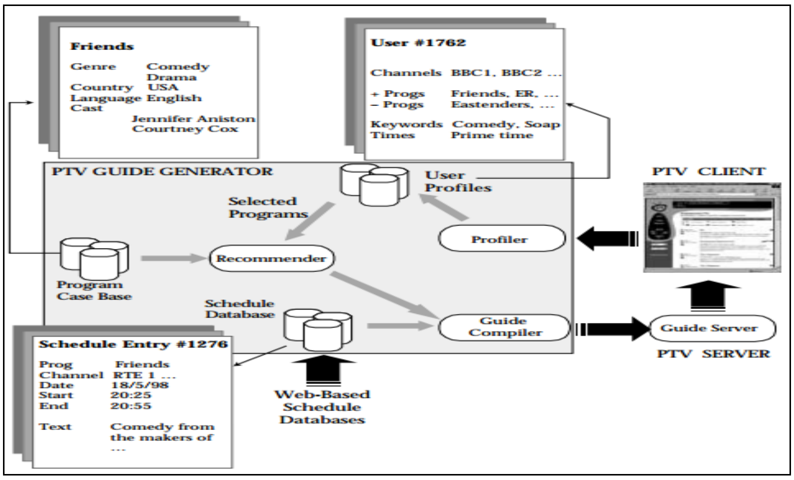

Figure 3: PTV System Architecture [25].

It employs content-based methods founded on textual descriptions of TV shows and collaborative information about the precedence of other users. Suggestions from the two methods are integrated in the final suggested program [25].

V. CONCLUSION

Accordingly, these days with technology improvement and also increasing the quantity of data we need a method and system that can help people to find their interests and their items with less effort and also with spending less time with more accurate. There are several ways that we can exploit them to reach these goals like Collaborative filtering (CF) that suggests items based on history valuation of all users communally, Content-base filtering which recommend according to previous users' precedence, and also Hybrid system that is combination of two techniques foresaid. These approaches have several advantages and disadvantages that at this research has tried to focus mostly on the recommendation approaches and their weaknesses. Although, recommendation systems with these conditions help users to find their preferences a lot they must be improved more and more.

REFERENCES

[1] Riedl, J., Jameson, A., & Konstan, J. (2004). "AI Techniques for Personalized Recommendation." Proposal.

[2] P. Resnick, H. R. Varian. (1997). "Recommender systems." Commun. ACM 40: 13.

[3] Melville, P., Sindhwani, V, "Recommender Systems." IBM T.J. Watson Research Center, Yorktown Heights, NY 10598: 21.

[4] Takacs.G, Pil´aszy.I, Nemeth.B, Tikk.D.(2009). "Scalable Collaborative Filtering Approaches for Large Recommender Systems." Journal of Machine Learning Research 10: 34.

[5] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry.(1992). "Using collaborative filtering to weave an information tapestry". Communications of the ACM, 10.

[6] Hanani.U, Shapira.B, Shoval.P, (2001), "Information Filtering: Overview of Issues, Research and Systems". User Modeling and User-Adapted Interaction, 57.

[7] Shardanand.U, Maes.P. (1995). "Social Information Filtering: Algorithms for Automating "Word of Mouth". In CHI '95: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA. 8.

[8] Aha.W, Kibler.D, Albert.M. (1991). "Instance-Based Learning Algorithms". Machine Learning, 30.

[9] Sarwar.B, Karypis.G, Konstan.J, and Riedl.J. (2001). "Item-Based Collaborative Filtering Recommendation Algorithms". In WWW '01: Proceedings of the 10th International Conference on World Wide Web, New York, NY, USA.11.

[10] Karypis.G. (2001). "Evaluation of Item-Based Top-N Recommendation Algorithms". In CIKM '01: Proceedings of the Tenth International Conference on Information and Knowledge Management, New York, NY, USA. 8.

[11] Herlocker.J, Konstan.J, Al Borchers, and Riedl.J. (1999). "An Algorithmic Framework for Performing Collaborative Filtering". In SIGIR '99: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA, 8.

[12] Breese. J. S, Heckerman. D, Kadie. C. (1998). "Empirical Analysis of Predictive Algorithms for Collaborative Filtering." Microsoft Research: 19.

[13] Chickering, D., Heckerman, D., and Meek, C. (1997). "A Bayesian approach to learning Bayesian networks with local structure".

[14] Burke, R. (2002). "Hybrid Recommender Systems: Survey and Experiments." User Modeling and User- Adapted Interaction: 40.

[15] Andrew I. Schein, A. P., Lyle H. Ungar, and David M. Pennock (2002). "Methods and Metrics for Cold-start Recommendations." In SIGIR '02: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,New York: 8.

[16] Mark Claypool, A. G., Tim Miranda, Pavel Murnikov, Dmitry Netes, and Matthew Sartin. (1999). "Combining Content-Based and Collaborative Filters in an Online Newspaper." In Proceedings of ACM SIGIR Workshop on Recommender Systems.

[17] Burke, R. (1999). "Integrating Knowledge-Based and Collaborative-Filtering Recommender Systems." In Proceedings of the AAAI Workshop on AI in Electronic Commerce: 4.

[18] Lang, K. (1995). "NewsWeeder: Learning to Filter Netnews." In ICML '95: Proceedings of the 12th International Conference on Machine Learning,San Mateo, CA, USA,: 9.

[19] Roy, R. J. M. a. L. (2000). "Content-Based Book Recommending Using Learning for Text Categorization." In DL '00: Proceedings of the Fifth ACM Conference on Digital Libraries, New York, NY: 10.

[20] Rashid.A.M, Albert.I,Cosley.D, K.Lam.S., McNee.S, A. Konstan.J,and Riedl.J.(2002). "Getting to Know You: Learning New User Preferences in Recommender Systems." In IUI '02: Proceedings of the 7th International Conference on Intelligent User Interfaces,New York, NY: 8.

[21] Pazzani, M. (1999). "A framework for collaborative, content-based and demographic filtering." Artificial Intelligence Review 13: 16.

[22] Balabanovi´c, M, Shoham, Y. (1997). "Fab: Content-based, collaborative recommendation." Communications of the ACM 40: 7.

[23] McNee, S., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S., Rashid, A., Konstan, J., and Riedl, J. (2002). "On the recommending of citations for research papers". In Proceedings of the 2002 ACM Conference on Computer-Supported Cooperative Work. ACM Press, New Orleans, LA, USA, 10.

[24] Torres, R., McNee, S., Abel, M., Konstan, J., and Riedl, J. (2004). "Enhancing digital libraries with TechLens". In Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries. ACM Press, Tuscon, AZ, USA, 9.

[25] Smyth, B., & Cotter, P. (2001). Personalized electronic program guides for digital TV. Ai Magazine, 22(2), 89.

[25] Nichols, D. M. "Implicit Rating and Filtering." 6.

[26] Oard, D.W. and Marchionini, G. (1996), "A Conceptual Framework for Text Filtering", Technical Report CARTR-830, Human Computer Interaction Laboratory, University of Maryland at College Park.