A Survey on Recommender Systems based on Collaborative Filtering Technique

Atisha Sachan
Department of Computer Science And Engineering
LNCT, BHOPAL, MP, INDIA

Vineet Richariya
Head Of The Department Of Computer Science And Engineering
LNCT, BHOPAL, MP, INDIA

Abstract- **Nowadays Product advertisement and viewer's choice are two important part of marketing. These two parts generate a system that system is known as Recommender system. Recommender system plays a vital role in internet technology for data gathering and rating up a data. There are four types of filtering technique used in Recommender System-demographic, content, collaborative and hybrid. The most widely and popularly used technique is collaborative filtering. In this paper we describe a little about former three techniques but mainly focuses on collaborative filtering, their types and their major challenges such as cold start problem, data sparsity, scalability, accuracy etc.**

Keywords – Recommender System, Collaborative Filtering, Cold Start, Sparsity, Accuracy.

I. INTRODUCTION

The explosive growth of e-commerce and online environments has made the issue of information search and selection increasingly serious; users are overloaded by options to consider and they may not have the time or knowledge to personally evaluate these options. Recommender systems[5] have proven to be a valuable way for online users to cope with the information overload and have become one of the most powerful and popular tools in electronic commerce. Correspondingly, various techniques for recommendation generation have been proposed. During the last decade, many of them have also been successfully deployed in commercial environments. Recommender form or work from a specific type of information filtering system technique that attempts to recommend information items (movies, TV program/show/episode, video on demand, music, books, news, images, web pages, scientific literature such as research papers etc.) or social elements (e.g. people, events or groups) that are likely to be of interest to the user. A recommender system helps users that have no sufficient competence to evaluate the, potentially overwhelming, number of alternatives. In their simplest form recommender systems provide a personalized and ranked lists of items by predicting what the most suitable items are, based on the user's history, preferences and constraints. Typically, a recommender system compares a user profile to some reference characteristics, and seeks to predict the 'rating' or 'preference' that a user would give to an item they had not yet considered. These characteristics may be from the information item (the content-based approach) or the user's social environment (the collaborative filtering approach).
Data collections are done by two methods-explicitly and implicitly.

Explicit data collections include the following:

- Asking a user to rate an item on a sliding scale.
- Asking a user to rank a collection of items from favourite to least favourite.
- Presenting two items to a user and asking him/her to choose the better one of them.
- Asking a user to create a list of items that he/she likes.

Implicit data collection includes the following:

- Observing the items that a user views in an online store.
- Analyzing item/user viewing time.
- Keeping a record of the items that a user purchases online.

- Obtaining a list of items that a user has listened to or watched on his/her computer.
- Analyzing the user's social network and discovering similar likes and dislikes.

II. TYPES OF FILTERING

2.1 Demographic Information Filtering–

Categorize users or items based on their personal attributes and make recommendation based on demographic categorizations. In other words we can say that in the filtering based on demographic information, users are classified by their features, and recommendation is given to the class of demographic information.

2.2 Content Based Filtering –

Another filtering that is widely used in recommender systems is content-based filtering. Content based filtering methods are based on the information about the items that are going to be recommended. In other words, these algorithms try to recommend the items similar to those that a user liked in the past. In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. This approach has its roots in information retrieval and information filtering research. Basically those methods use an item profile i.e. a set of attributes (features) characterizing the item within the system. The system creates a content based profile of users based on a weighted vector of item features. The weights denote the importance of each feature to the user and can be computed from individually rated content vectors using a variety of techniques. Simple approaches use the average values of the rated item vector while other sophisticated methods use Bayesian Classifiers (and other machine learning techniques, including clustering, decision trees, and artificial neural networks) in order to estimate the probability that the user is going to like the item.

Content based system has features such as simplicity and effectiveness, but also some drawbacks:

- It is difficult to distinguish the quality of the filtering results from the same subject. Since the quantity of information increases rapidly, the information of the same subject increases too, making the efficiency and quality of the content-based system much reduced in a long term.

- Incapable to discover the information of user’s new interests, the system could only locate the information similar to user’s current interests.

2.3 Collaborative Filtering-

Another filtering technology that is widely used in recommender systems is Collaborative Filtering. Compared with the content-based filtering system, collaborative filtering system could automatically filter the information that the system could not analyze and represent, and recommend up-to-date information. Collaborative filtering methods are based on collecting and analyzing a large amount of information on users’ behaviour, activity or preferences and predicting what users will like based on their similarity to other users. One of the most common types of Collaborative Filtering is item-to-item collaborative filtering (people who buy x also buy y), an algorithm popularized by Amazon.com recommender system. User-based collaborative filtering attempts to model the social process of asking a friend for a recommendation. A particular type of collaborative filtering algorithms uses matrix factorization, a low-rank matrix approximation technique. A key advantage of the collaborative filtering approach is that it does not rely on machine analyzable content and therefore it is capable of accurately recommending complex items such as movies without requiring an "understanding" of the item itself.

Many collaborative filtering techniques have been developed. They can be categorized into two types:

- *<u>Memory Based Collaborative Filtering:</u>*
  Memory-based CF uses user-to-user or item-to-item correlations based on users’ rating behaviour to recommend or predict ratings for users on future items. Correlations can be measured by various distance metrics, such as Pearson correlation coefficient, cosine distance, and Euclidean distance. Memory-based collaborative filtering uses the whole training set each time it computes a prediction, which makes it easy to incorporate new data but suffers slow performance on large data sets. Speedup can be achieved by pre-calculating correlations and other needed information and incrementally updating them. For some

applications, however, the size requirement makes the approach infeasible. It can perform with high recommendation accuracy, and new data can also be easily applied into recommendation. However, it is costly in computing and with bad scalability.

- *Model Based Collaborative Filtering:*

Unlike memory-based CF, model-based approach does not use the whole data set to compute a prediction. Instead, it builds a model of the data based on a training set and uses that model to predict future ratings. For example, clustering based CF method builds a model of the data set as clusters of users, and then uses the ratings of users within the cluster to predict. A very successful model-based method is the Singular Value Decomposition (SVD) [3], which represents the data by a set of vectors, one for each item and user, such that the dot product of the user vector and the movie vector is the best approximation for the training set. Typical the model building process is computationally expensive and memory intensive. After models are constructed, predictions can be done very fast with small memory requirement. Model-based CF methods usually achieve less accurate prediction than memory-based methods on dense data sets where a large fraction of user-item values are available in the training set, but perform better on sparse data sets.

III. CHALLENGES OF COLLABORATIVE FILTERING

There are some potential problems with the Collaborative Filtering RS. One is the scalability, which is how quickly a recommender system can generate recommendation; second one is sparsity and also cold start Problem and better accuracy. We discuss all of them below[17]:

Scalability:

In many of the environment that these systems make recommendation in, there are millions of users and products. Thus a large amount of computation power is often necessary to calculate recommendation. For example, with tens of millions of customers (*M*) and millions of distinct catalog items (*N*), a *CF* algorithm with the complexity of $O(n)$ is already too large. As well, many systems need to react immediately to online requirements and make recommendations for all users regardless of their purchases and ratings history, which demands a high scalability of a *CF* system [3].

Data Sparsity:

The number of items sold on major e commerce sites is extremely large. The most active users will only have rated a small subset of the overall database. Thus even the most popular items have very few ratings.

Cold Start Problem:

The *cold start* problem occurs when a new user or item has just entered the system, it is difficult to find similar ones because there is not enough information (in some literature, the *cold start* problem is also called the *new user problem* or *new item problem* [18, 19]). New items cannot be recommended until some users rate it, and new users are unlikely given good recommendations because of the lack of their rating or purchase history.

IV. RELATED SURVEY

We study various research paper and journal and know about recommender systems, collaborative filtering also their drawbacks. All methodology and process are not described here. But some related work in the field of filtering in Recommender System discuss by the name of authors and their respective title.

A. By Guangping Zhuo, Jingyu Sun and Xueli Yu[8] **"A Framework for Multi-Type Recommendations"** deals in the field of web minning concern on some drawbacks in collaborative filtering and also on multi type recommendation. Collaborative filtering (CF) is an effective method of recommender systems (RS) has been widely used in online stores. However, CF suffers some weaknesses: problems with new users (cold start), data sparseness, difficulty in spotting "malicious" or "unreliable" users and so on. Additionally CF can't recommend different type items at the same time. So in order to make it adaptive new Web applications, such as urban computing, visit schedule planning and so on, introduced a new recommendation framework, which combines CF and case-based reasoning (CBR) to improve performance of RS. Based on this framework, the authors have developed a semantic search demo system—MyVisit, which shows that our proposed framework is an effective recommendation model. Two key algorithms, **MIFA** and **RAA**, are used. Additionally, authors have validated them using an application instance, which is a demo system for recommending multi type recommendations combining CF and CBR.

Advantage of this method is that it involves a few of cases in the online and adjusts the rating of main items through associative other type items in order to find fit recommendations.

B. By Yechun Jiang, Jianxun Liu, Mingdong Tang and Yechun Jiang, Jianxun Liu, Mingdong Tang [9**] "An Effective Web Service Recommendation Method based on Personalized Collaborative Filtering"**. Describing an effective personalized collaborative filtering method for Web service recommendation. A key component of Web service recommendation techniques is computation of similarity measurement of Web services. Different from the Pearson Correlation Coefficient (PCC) similarity measurement, they take into account the personalized influence of services when computing similarity measurement between users and personalized influence of services. Based on the similarity measurement model of Web services, develop an effective Personalized Hybrid Collaborative Filtering (PHCF) technique by integrating personalized user-based algorithm and personalized item-based algorithm. Also conduct series of experiments based on real Web service QoS dataset WSRec [20] which contains more than 1.5 millions test results of 150 service users in different countries on 100 publicly available Web services located all over the world. Experimental results show that the method improves accuracy of recommendation of Web services significantly.

C. By Qian Wang, Xianhu Yuan, Min Sun [1**] "Collaborative Filtering Recommendation Algorithm based on Hybrid User Model".** Collaborative filtering faces challenges of scalability and also recommendation accuracy so the paper proposes a hybrid user model to remove some of its drawbacks. The recommender system based on this model not only holds the advantage of recommendation accuracy in memory-based method, but also has the scalability as good as model-based method. The user model is constructed based on item combination feature and demographic information, and it focuses on searching for set of neighbouring users shared with same interest, which helps to improve system scalability. To enhance recommendation accuracy, each feature in user model is given a different weight when computing the similarity between users. Genetic algorithm is adopted to learn the weight values of features. Methodology proposed improves recommendation accuracy and scalability to a certain extent. It constructs a concise and representative hybrid user model, and combines and integrates item ratings, item detailed description and demographic information together, which raises the density of data and improves the problem of sparse data. Besides, genetic algorithm is adopted to learn a best feature weight vector in computation of the nearest neighbor set, which helps to get a more accurate similarity. The experiment shows that algorithm proposed in the paper can get a recommendation with higher accuracy, compared to methods of TCF and CCCF.

D. By Chuangguang Huang and Jian Yin [10] **"Effective Association Clusters Filtering to Cold-Start Recommendations".** The paper focuses on how to overcome cold-start problem in the traditional research of Recommendations System (RS). The popular technique of RS is Collaborative Filtering (CF). While in real online RS, CF can't practically solve cold-start problem for the sparsity ratings dataset. The paper proposed a novel efficiently association clusters filtering (ACF) algorithm. Considering hybrid approaches, using clustering and also filtering to relieve cold-start problem. ACF algorithm establishes clusters models based on the ratings matrix. We assume the users in the same cluster; they will have the same interests. On the other hand, different users in different clusters present they will have less common interests. The more users ratings for some item in the cluster, can delegate the opinion of the cluster. So we can use the opinion of the cluster to predict the unknown ratings. Throughout the experiments, our method can enlarge the prediction scope and improve the accuracy. The advantage of our algorithm is clearer if we use the more sparsity dataset. We see our main contribution as a detailed study of a number of different clusters filter generation methods and demonstration that a very few number of simple association clusters help algorithms work better in cold-start situation, with negligible impact on non-cold-start recommendation accuracy and system efficiency.

E. By Mustansar Ali Ghazanfar and Adam Prugel-Bennett [11] **"A Scalable, Accurate Hybrid Recommender System".** The paper proposes a unique cascading hybrid recommendation approach by combining the rating, feature, and demographic information about items. They empirically show that their approach outperforms the state of the art recommender system algorithms, and eliminates recorded problems with recommender systems. Since there are three main types of recommender systems: collaborative filtering, content-based filtering, and demographic recommender systems. Collaborative filtering recommender systems recommend items by taking into account the taste (in terms of preferences of items) of users, under the assumption that users will be interested in items that users similar to them have rated highly. Content-based filtering recommender systems recommend items based on the textual information of an item, under the assumption that users will like similar items to the ones they liked before. Demographic recommender systems categorize users or items based on their personal attribute and make recommendation based on demographic categorizations. These systems suffer from scalability, data sparsity, and cold-start problems resulting in poor quality recommendations and reduced coverage. So they combine all these filtering to form a hybrid recommender system.

F. By Liang He and Faqing Wu [12] **"A Time-context-based Collaborative Filtering Algorithm".** The paper incorporates the time-context, one of the most important contexts, into the traditional collaborative filtering algorithm and proposes a Time context-Based Collaborative Filtering (TBCF) Algorithm to improve the performance for traditional collaborative filtering algorithm. Experiments evaluating this approach are carried out on real dataset taken from movie recommendation system provided by MovieLens web site. The result shows the proposed approach can improve predication accuracy and recall ratio compared with existing methods. The time context is a very important factor in recommendation system. And the paper introduced time interval into the traditional user-based collaborative filtering algorithm. The strategies proposed improved both the prediction accuracy and recall ratio of standard user-based collaborative filtering methods.

G. By Ling Yun, Wang Xun and Gu Huamao [13**] "A Hybrid Information Filtering Algorithm Based on Distributed Web log Mining"**. For distributed large commercial mirror sites, the paper presents a hybrid information filtering algorithm based on distributed web log mining. Based on multiagent technology, the algorithm pre-processes the web logs of mirror sites, in which the web page's manual rating is replaced by user browsing preference, and then user access matrix is constructed and standardized.
The paper proposes a distributed web log mining based hybrid filtering algorithm. To solve the problem that users are reluctant to rate web pages, this paper establishes the user access matrix on the basis of web log mining to gather fundamental data for both filtering. For the sparseness of user rating data of collaborative filtering, a collaborative filtering algorithm is proposed based on web page rating prediction, which effectively overcomes the drawbacks of traditional similarity measuring methods under circumstances of data sparseness and improves the accuracy of target user's calculation of the nearest neighbor. To address the drawbacks of those two filtering models, the paper presents a hybrid filtering model. With the optimal weight, this model further improves the recommendation quality. But this algorithm is tested only in the simulation, thus lacking the test under distributed net environment, so the reliability and performance of the algorithm needs to be further proved.

H. By Ibrahim A. Almosallam and Yi Shang [8] **"A New Adaptive Framework for Collaborative Filtering Prediction".** The paper focused on memory-based collaborative filtering (CF). Existing CF techniques work well on dense data but poorly on sparse data. To address this weakness, the paper proposed to use z-scores instead of explicit ratings and introduce a mechanism that adaptively combines global statistics with item-based values based on data density level. They present a new adaptive framework that encapsulates various CF algorithms and the relationships among them. An adaptive CF predictor is developed that can self adapt from user-based to item-based to hybrid methods based on the amount of available ratings. The experimental results show that the new predictor consistently obtained more accurate predictions than existing CF methods, with the most significant improvement on sparse data sets. When applied to the Netflix Challenge data set, our method performed better than existing CF and singular value decomposition (SVD) methods and achieved 4.67% improvement over Netflix's system.

I. By Cane Wing-ki Leung, Stephen Chi-fai Chan and Fu-lai Chung [15] **"Applying Cross-Level Association Rule Mining to Cold-Start Recommendations".** The paper proposed a novel hybrid recommendation algorithm for addressing the well-known cold-start problem in Collaborative Filtering (CF). The algorithm makes use of Cross-Level Association RulEs (CLARE) to integrate content information about domain items into collaborative filters. They first introduce a preference model comprising user-item and item-item relationships, and described the CLARE algorithm for generating cold-start recommendations. When no recommendations can be generated for an item from ratings data, CLARE takes into consideration the attributes of the item for generating cold-start recommendations. Experimental results validated the ability of CLARE to recommend cold-start items and to improve significantly the number of recommendable items in a system. They experimented with only one type of attribute (cast) for mining CARs as an initial effort. They studied the behaviour of CLARE using more attribute types with varying characteristics, and obtained improved recommendation quality and coverage.

J. By Leo Iaquinta, Anna Lisa Gentile, Pasquale Lops, Marco de Gemmis and Giovanni Semeraro [7] **"A Hybrid Content-Collaborative Recommender System Integrated into an Electronic Performance Support System".** The paper proposed the adoption in an EPSS of a novel hybrid recommender that implements a neighbourhood formation process based on the idea of grouping users by computing similarities between their semantic user profiles instead of their rating style. Our hybrid recommender overcomes some shortcomings of pure CF systems:

- ***Sparsity Problem*** - interpreted the MAE improvement as a direct consequence of the proposed neighbourhood formation strategy and this Improvement is particularly evident in case of data sparsity, when the strategy based on the Pearson's correlation coefficient is more likely to fail.
- ***Lack of Transparency Problem*** - the adoption of synset-based profiles to select the neighbourhood of users gives the possibility to understand why some users have been selected for producing recommendations. Profiles are represented by senses instead of words, thus a certain level of system transparency has been added. To the best of our knowledge, the clustering of synset-based profiles for the process of neighbourhood selection is a novel contribution in the area of CF systems. The scenario of the experimental evaluation was different from the JUMP project domain: it simply represents a proof of concept in order to verify the quality of the hybrid recommender. As future work, we expect to integrate the hybrid recommender into the JUMP EPSS and to run an experimental evaluation.

K. By Manos Papagelis, Dimitris Plexousakis, Themistoklis Kutsuras **"Alleviating the Sparsity Problem of Collaborative Filtering Using Trust Inferences"**. In this research, our main objective was to describe a method that is able to provide high-quality recommendations even when information available is insufficient. Our work employs theoretical results of research conducted in areas of social networks and trust management in order to develop a computational trust model for recommendation systems. To deal with the sparsity problem we proposed a method that is based on trust inferences. Trust inferences are transitive associations between users that participate in the underlying social network. Employment of this model provides additional information to Collaborative Filtering algorithm and remarkably relaxes the sparsity and the cold-start problems. Furthermore, our model considers the subjective notion of trust and reflects the way in which it is raised in real world social networks. Subjectiveness is defined in terms of confidence and uncertainty properties that are applied to the network associations. We have experimentally evaluated our method according to the impact that trust inferences have to sparsity and according to recommendation quality. Our experimental results indicate that our method succeeds in providing additional information to the Collaborative Filtering algorithm while it outperforms the quality performance of the classic CF method. The methodology described is general and may probably be easily adopted to alleviate the sparsity problem in other application areas, especially where underlying social networks can be identified.

V. CONCLUSION

We study various research papers including upper ones and results on this fact that Collaborative Filtering is mostly used filtering technique but then also it has some issues related to sparsity, accuracy, scalability etc. There have been many researches and also results given by many authors. They all are focuses on Scalability, Cold Start, Sparsity and Accuracy. But there is not much work was done on sparsity issue. Since today internet data is growing fastly thats why sparsity also increases as new records, items, things, music, data etc are increasing and loaded day by day. In future work we research on sparsity issue as it is also the important challenge that recommender system faces today and in future also.

REFERENCES

[1] Qian Wang, Xianhu Yuan, Min Sun "Collaborative Filtering Recommendation Algorithm based on Hybrid User Model", FSKD, 2010.
[2] Jong Seo Lee "Survey of Recommender Systems (Collaborative Filtering)"
[3] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, 2003.
[4] MovieLens data, http://www.grouplens.org/.
[5] http://www.google.com/
[6] http://www.googlescholar.com/
[7] Leo Iaquinta, Anna Lisa Gentile, Pasquale Lops, Marco de Gemmis and Giovanni Semeraro "A Hybrid Content-Collaborative Recommender System Integrated into an Electronic Performance Support System", Seventh International Conference on Hybrid Intelligent Systems, 2007.
[8] Guangping Zhuo, Jingyu Sun and Xueli Yu "A Framework for Multi-Type Recommendations", Eighth International Conference on Fuzzy Systems and Knowledge Discovery, 2007.
[9] Yechun Jiang, Jianxun Liu, Mingdong Tang and Xiaoqing (Frank) Liu "An Effective Web Service Recommendation Method based on Personalized Collaborative Filtering", 2011 IEEE International Conference on Web Services.
[10] Chuangguang Huang and Jian Yin "Effective Association Clusters Filtering to Cold-Start Recommendations", 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery.
[11] Mustansar Ali Ghazanfar and Adam Prugel-Bennett," A Scalable, Accurate Hybrid Recommender System", 2010 Third International Conference on Knowledge Discovery and Data Mining.
[12] Liang He and Faqing Wu, "A Time-context-based Collaborative Filtering Algorithm".

[13] Ling Yun, Wang Xun and Gu Huamao "A Hybrid Information Filtering Algorithm Based on Distributed Web log Mining", Third 2008 International Conference on Convergence and Hybrid Information Technology.
[14] Ibrahim A. Almosallam and Yi Shang "A New Adaptive Framework for Collaborative Filtering Prediction", *2008 IEEE Congress on Evolutionary Computation (CEC 2008).*
[15] Cane Wing-ki Leung, Stephen Chi-fai Chan and Fu-lai Chung "Applying Cross-Level Association Rule Mining to Cold-Start Recommendations", 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.
[16] Manos Papagelis, Dimitris Plexousakis and Themistoklis Kutsuras "Alleviating the Sparsity Problem of Collaborative Filtering Using Trust Inferences".
[17] Xiaoyuan Su and Taghi M. Khoshgoftaar "A Survey of Collaborative Filtering Techniques", Hindawi Publishing Corporation, Advances in Artificial Intelligence Volume 2009.
[18] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005.
[19] K. Yu, A. Schwaighofer, V. Tresp, X. Xu, and H.-P. Kriegel, "Probabilistic memory-based collaborative filtering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 1, pp. 56–69, 2004.
[20] Z.B. Zheng, H. Ma, M.R. Lyu, and I. King, "WSRec: a collaborative filtering based Web service recommendation system," Proc. 7th International Conference on Web Services (ICWS 2009), 2009, pp. 437-444.