Collaborative Filtering Approach Based Recommender Systems

R.N.Ravikumar [1] and S.Aarthi [2]

[1]Department of Computer Science and Engineering
[1]VMKV Engineering College,Vinayaka Missions University, India.
[1]

[2]Department of Computer Science
[2]PKR Arts College for Women, Bharatiyar University, India.
[2]

Abstract

One of the potent personalization technologies powering the adaptive web is collaborative filtering. Collaborative filtering (CF) is the process of filtering or evaluating items through the opinions of other people. CF technology brings together the opinions of large interconnected communities on the web, supporting filtering of substantial quantities of data. In this chapter we introduce the core concepts of collaborative filtering, its primary uses for users of the adaptive web, the theory and practice of CF algorithms, and design decisions regarding rating systems and acquisition of ratings. We also discuss how to evaluate CF systems, and the evolution of rich interaction interfaces. We close the chapter with discussions of the challenges of privacy particular to a CF recommendation service and important open research questions in the field.

Index terms-- *Collaborative filtering, Recommendation.*

I. INTRODUCTION

Collaborative Filtering is the process of filtering or evaluating items using the opinions of other people. While the term collaborative filtering (CF) has only been around for a little more than a decade, CF takes its roots from something humans have been doing for centuries sharing opinions with others [1]. For years, people have stood over the back fence or in the office break room and discussed books they have read, restaurants they have tried, and music they have seen then used these discussions to form opinions. Computers and the web allow us to advance beyond simple word-of-mouth. Instead of limiting ourselves to tens or hundreds of individuals the Internet allows us to consider the opinions of thousands. The speed of computers allows us to process these opinions in real time and determine not only what a much larger community thinks of an item, but also develop a truly personalized view of that item using the opinions most appropriate for a given user or group of users.

The term user refers to any individual who provides ratings to a system. Most often, we use this term to refer to the people using a system to receive information (e.g., recommendations) although it also refers to those who provided the data (ratings) used in generating this information. Collaborative filtering systems produce predictions or recommendations for a given user and one or more items. Items can consist of anything for which a

human can provide a rating, such as art, books, CDs, journal articles, or vacation destinations. Ratings in a collaborative filtering system can take on a variety of forms.

- ❖ Scalar ratings can consist of either numerical ratings, such as the 1-5 stars provided in ordinal ratings such as strongly agree, agree, neutral, disagree, Strongly disagree.
- ❖ Binary ratings model choices between agree/disagree or good/bad.
- ❖ Unary ratings can indicate that a user has observed or purchased an item, or other-wise rated the item positively. The absence of a rating indicates that we have no in-formation relating the user to the item.
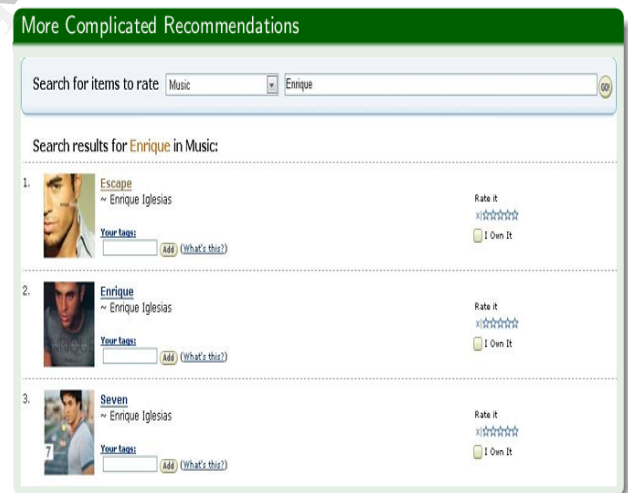

Fig.1 Collaborative filtering to predict that this user is likely to rate the music "Holes" 4 out of 5 stars.

Ratings may be gathered through explicit means, implicit means, or both. Explicit ratings are those where a user is asked to provide an opinion on an item. Implicit ratings are those inferred from a user's actions. For example, a user who visits a product page perhaps has some interest in that product while a user who subsequently purchases the product may have a much stronger interest in that product.

II. COLLABORATIVE FILTERING AND THE ADAPTIVE WEB

These early collaborative filtering systems were designed to explicitly provide users with information about items. That is, users visited a website for the purpose of receiving recommendations from the CF system. Later, websites began to use CF systems behind the scenes to adapt their content to users, such as choosing which news articles a website should be presenting prominently to a user. Providers of information on the web must deal with limited user attention and limited screen space. Collaborative filtering can predict what information users are likely to want to see, enabling providers to select subsets of information to display in the limited screen space. By placing that information prominently, it enables the user to Maximize their limited attention. In this way, collaborative filtering enables the web to adapt to each individual user's needs.

III. RELATED WORKS

3.1 Recommend Items

Show a list of items to a user, in order of how useful they might be. Often this is described as predicting what the user would rate the item, then ranking the items by this predicted rating [2]. However, some successful recommendation algorithms do not compute predicted rating values at all. For example, Amazon's recommendation algorithm aggregates items similar to a user's purchases and ratings without ever computing a predicted rating. Instead of displaying a personalized predicted rating, their user interface displays the average customer rating [3]. As a result, the recommendation list may appear out of order with respect to the displayed average rating value. In many applications, picking the top few items well is crucial; producing predicted values is secondary.

3.2 Predict For a Given Item

Given a particular item, calculate its predicted rating. Note that prediction can be more demanding than recommendation. To recommend items, a system only needs to be prepared to offer a few alternatives, but not all. Some algorithms take advantage of this to be more scalable by saving memory and computation time [7]. To provide predictions for a particular item, a system must be prepared to say something about any requested item, even rarely rated ones. How does a system decide how a particular user would rate a requested item if very few users let alone users similar to the particular user have rated the item? Personalized predictions may be challenging, if not impossible.

3.3 Constrained Recommendations

Recommend from a set of items. Given a particular set or a constraint that gives a set of items, recommend from within that set. For example: "Consider the following scenario. Mary's 8-year-old nephew is visiting for the weekend, and she would like to take him to the music. She would like a comedy or family music rated no "higher" than PG-13. She would prefer that the music contain no sex, violence or offensive language, last less than two hours and, if possible, show at a theater in her neighborhood. Finally, she would like to select music that she herself might enjoy." propose a "meta-recommendation system" that generates recommendations from a blending of multiple recommendation sources. Users define preferences and requirements through a web form that restricts the set of potential candidate items. Recommendations are based on a ranking of how well the items within this set match the provided preferences, SQL-like language as a desired extension in a "next-generation" recommendation system Such a system might accept queries such as "RECOMMEND Movie TO User BASED ON Rating FROM Movie Recommender WHERE Movie.Length < 120 AND Movie.Rating < 3 AND User.City = Movie.Location"[9].

IV. COLLABORATIVE FILTERING APPROACH

Collaborative filtering is a technique that automatically predicts the interest of active users by collecting rating information from other similar users or items. Collaborative filtering approaches in which neighborhood based approach is most widely used. Neighborhood collaborative filtering includes two type of approach.

1. User-based approach
2. Item-based approach

User-based approaches predict the rating of active users based on the ratings of their similar users and Item-based approaches predict the rating of active users based on the computed information of items similar to those chosen by the active user. Both the User based and item-based approaches often use the Pearson Correlation Coefficient algorithm (PCC) as the similarity computation methods. This PCC based collaborative filtering generally can achieve higher performance than any other popular algorithm because this considers the differences of the user rating style. In web the rating data are always unavailable since the information on web is less structured and more diverse. There are different methods which all focus the user-item rating matrix using low-rank approximations which can be used to make further prediction. He premise behind these low dimensional factor models is that there are only a small number of factors influencing preferences, and that a user preference vector is determined by how each factor applies to that user. Here the query suggestion algorithms cannot be applied directly to most of the recommendation tasks on the web like query suggestion and image recommendation [2].

V. COLLABORATIVE FILTERING BASED RECOMMENDER SYSTEMS

Recommender systems have been evaluated in many, often incomparable, ways. In this article, we review the key decisions in evaluating collaborative filtering recommender systems: the user tasks being evaluated, the types of analysis and datasets being used, the ways in which

prediction quality is measured, the evaluation of prediction attributes other than quality, and the user-based evaluation of the system as a whole. In addition to reviewing the evaluation strategies used by prior researchers, we present empirical results from the analysis of various accuracy metrics on one con-tent domain where all the tested metrics collapsed roughly into three equivalence classes. Metrics Within each equivalency class were strongly correlated, while metrics from different equivalency classes were uncorrelated. Recommender systems use the opinions of a community of users to help individuals in that community more effectively identify content of interest from a potentially overwhelming set of choices [8]. One of the most successful technologies for recommender systems, called collaborative filtering, has been developed and improved over the past decade to the point where a wide variety of algorithms exist for generating recommendations and additional qualitative evaluation techniques.
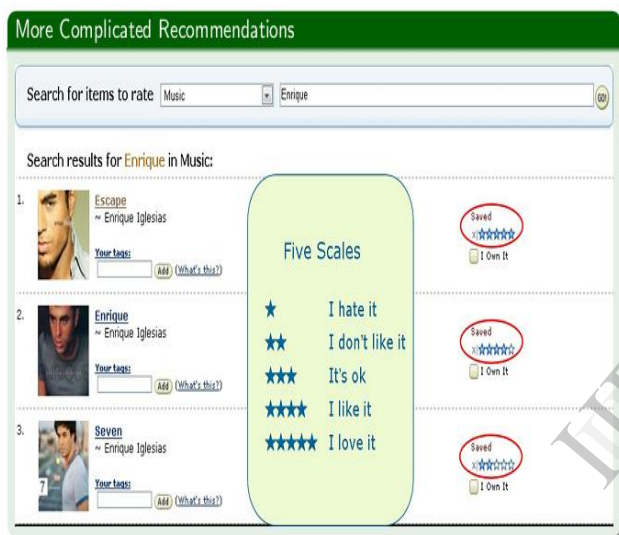

Fig.2 Ratings that strongly influenced a particular recommendation

VI. TECHNIQUES IN RECOMMENDER SYSTEM

Recommender systems apply data mining methods and prediction algorithms to predict users' interest on information, products and services among the marvelous amount of available items. The central component of all recommender systems is the user prototypical that contains information about the individual preferences which control his or her behavior in a complex environment of web-based systems.

6.1 Apriority Algorithm

The recommendation system using apriori algorithms is a classic algorithm for learning association rules [8]. Apriori is designed to function on databases containing transactions. Additional algorithms are calculated for finding association rules in data having no relations. It is collective in association rule mining, given a set of element sets, the algorithm attempts to find subsets which are common to at least a lowest number C of the item groups. Apriority is a bottom up methodology, where

collective subsets are extended one item at a time and sets of candidates are verified against the data. The algorithm trimmings when no further successful extensions are found.

6.2 Collaborative Filtering

The aim of a collaborative filtering algorithm is to propose new items or to calculate the utility of a certain item for a particular user based on the user's prevision likings and the opinions of other like-minded users [1].

6.2.1 Item-Based Collaborative Filtering

A predictable a different method in the area of filtering algorithms, that was proposed newly is based on item relations and not on user relations, as in typical Collaborative Filtering. In Item-based Collaborative Clarifying process, we look into the group of items, that the dynamic user, has rated, compute how like they are to the goal item and then choice the k most similar items {i1, i2, ..., ik}, based on their parallel similarities {si1, si2, ..., sik}. The calculations can then be calculated by taking a weighted average of the dynamic user's scores on these associated items. The first step in this new approach is the Representation. Its resolve is the related as with the classic Collaborative Filtering procedure: represent the data in an ordered manner [3].

6.2.2 Content-Based Collaborative Filtering

The basic idea behind Content-Boosted Collaborative Filtering is to use a content-based predictor to enhance current user data, communicated via the user-item matrix, R, and then deliver personalized suggestions through collaborative filtering. The content-based analyst is practical on each row from the first user-item matrix, corresponding to every separate user, and gradually makes a pseudo user-item matrix, PR. At the end, each row, i, of the pseudo user-item matrix PR consists of the scores providing by user ui, when available, and those grades predicted by the content-based predictor [7].

6.3 Link Analysis Algorithm

New recommendation algorithm which we lately developed based on the thoughts from link analysis research [10]. Association analysis procedures have found essential application in Web page ranking and social network study.

VII. EVALUATION

Evaluation measures how well a collaborative filtering system is meeting its goals, either in absolute terms or in relation to alternative CF systems. Unfortunately, there is no well-accepted metric that can evaluate all-important criteria related to the performance of a CF system. The appropriate metric to choose may depend on the type of items being recommended, the user tasks supported by the CF system, and any external goals that the service providers may have (e.g., promotional or inventory depletion). An in depth discussion of evaluation considerations of collaboration filtering systems can be found in this section, we first discuss accuracy, which is generally considered the most important criteria to evaluate, and then briefly deal with some of the other criteria that may be important to evaluate and their associated metrics.

VIII. ONGOING CHALLENGES TO COLLABORATIVE FILTERING

8.1 Privacy and Security

In order to provide personalized information to users, CF systems need to know things about those users. In fact, the more the system knows about a user, the better predictions it can provide to that user. With this increased information stored by a system often comes an increased concern on the part of the user regarding what in-formation is collected, where and how it is stored, and how it is used. In centralized CF architectures, a single repository stores all user ratings. If the central server becomes compromised or corrupt, a user's anonymity can be destroyed. Users must trust that the CF provider will not use their preferences except for providing ratings and recommendations. Distributed architectures may deploy ratings or models to each user, risking exposure of information to every peer. To protect against this, researchers have developed security techniques building on encryption and shared keys. In these schemes, a user can encrypt their ratings, and peers can tally encrypted ratings. Once ratings are totaled, distributed agents use shared keys to decrypt the rating tallies, without being able to see the original ratings. Even systems that maintain the security of their users' ratings can be exploited to reveal personal information, particularly for users with unusual tastes and are most susceptible to exploitation. Unfortunately, it is often these esoteric users that are most valuable to recommender systems, because they can provide users with unexpectedly novel recommendations.

IX. CONCLUSION

Collaborative filtering is one of the core technologies that will power the adaptive web. Content-based personalization can be effective in limited circumstances, but for the most part, it will likely be decades or longer before our hardware and software technology can begin to automatically recognize the subtleties of information that are important to people particularly aspects of aesthetic taste. Until then, in order to filter information based on such complex dimensions, we need to include people in the loop, who analyze the information and condense their opinions into data that can be easily processed by software ratings. In this chapter, we have attempted to provide a snapshot of the current understanding of collaborative filtering systems and methods. By necessity, as masses of information become ubiquitously available, collaborative filtering will also become ubiquitous. In the process, we will continue to gain a deeper understanding of the dynamics of collaborative filtering.

REFERENCES

[1] Canny, J.: Collaborative Filtering with Privacy via Factor Analysis. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval. (2002) Tampere, Finland. ACM Press p. 238-245.

[2] Herlocker, J., Konstan, J.A., Terveen, L.G., Reidl, J.: Evaluating Collaborative Filtering Recommender Systems. ACM Transactions on Information Systems, (2004) 22(1): p. 5-53.

[3] Linden, G., Smith, B., York, J.: Amazon.Com Recommendations: Item-To-Item Collaborative Filtering. Internet Computing,IEEE,2003.7(1): p.76 80.

[4] Ramakrishnan, N., Keller, B.K., Mirza, B.J.: Privacy Risks in Recommender Systems. IEEE Internet Computing. 2001. p. 54-62.

[5] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: Grouplens: An Open Architecture For Collaborative Filtering Of Netnews. In Proceedings of the 1994 ACM conference on Computer supported cooperative work. (1994) Chapel Hill, North Carolina. ACM Press p. 175-186.

[6] Sarwar, B., Karypis, G., Konstan, J.A., Riedl, J.: Item-Based Collaborative Filtering Recommendation Algorithms. Proceedings of the 10th international conference on World Wide Web. (2001) Hong Kong. ACM Press p. 285-295.

[7] R. Agrawal and T. Imielinski ,"A. Swami: Mining Association Rules Between Sets of Items in Large Databases", SIGMOD Conference, pages 207-216, June, 1993.

[8] R. Agrawal and R. Srikant. "Fast algorithms for mining association rules", Proceedings of the 20th International Conference on Very Large Data Bases, Pages 487 - 499 Santiago, Chile, Sep,1994.

[9] Bardul M. Sarwar, George Karypis, Joseph A. Konstan, and John T. Riedl, "Item-based collaborative filtering recommendation algorithms," Proceedings of the 10th international conference on World Wide Web, Pages 285-295, Hong Kong, 2001.

[10] George Karypis, "Evaluation of item-based top-n recommendation algorithms," Proceedings of the tenth international conference on Information and knowledge management, Pages 247 - 254, 2001.

Ravikumar R N received his M.E Master of Engineering Degree in Computer Science and Engineering, from Vinayaka Missions University, Salem, India, in 2013. He received his B.E Bachelor of Engineering degree in Computer Science and Engineering, from Jayam College of Engineering and Technology, Anna University, India, in 2010. His research interests include data mining and its technologies, algorithms and applications, information retrieval, recommender systems.

Aarthi S received her BCA Bachelor of Computer Applications degree, from PKR Arts College for Women, Erode, India, in 2009. She received her MCA Master of Computer Applications degree, from Coimbatore Institute of Management and Technology, Coimbatore, India, in 2012. Currently, she is an M.Phil research scholar, from Department of Computer Science, at PKR Arts College for Women, Bharatiyar University, India. Her research interests include information retrieval, data mining, machine learning, recommender systems, and social network analysis.