A Survey on Various Approaches Used for Collaborative Filtering Based Recommendation

1Saudagar L. Jadhav, 2M. P. Mali

1,2Department of Computer Engineering, Vishwakarma Institute of Information Technology, Savitribai Phule Pune University, India
Email: 1saudagarjadhav@gmail.com, 2manisha.mali@viit.ac.in

Abstract— The explosive increase in the services or products put forth the information searching and selection as critical issue. Recommender Systems (RSs) are the most valuable tool useful to deal with such issues. RSs are the promising technology which helps user to choose the ideal services or products. RSs alludes services or products which user may be best-liked. These RSs are mainly based on four types of filtering Techniques-Demographic, Content-based, Collaborative and Hybird. Among these four, Collaborative Filtering (CF) finds the most successful technique in RS and hence widely used by many applications which used such RS. But this CF technique creates some challenges to the RSs. To overcome these challenges is becoming a crucial task. This paper mainly focuses on the challenges faced by RSs and how the various approaches are used with CF to overcome these challenges of RSs generated due to CF technique.

Keywords— Recommender System, Collaborative Filtering, Challenges

I. INTRODUCTION

Day by day the popularity of the Internet and web technology is increasing rapidly. 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 [1]. In this era of information overload, recommender systems play a dominant role in discovering valuable and interesting information for users searching among massively large databases. It is difficult to achieve what users would like because users sometimes do not realize what they want to search for [2]. A recommender system helps users that have no sufficient proficiency to guesstimate the imaginably overwhelming, number of alternatives. Recommender systems provide personalized recommendations to users, directing them to ideal, needed and interesting items. In simple word, the RSs provide a personalized and ranked lists of items by predicting what the most pertinent items are, based on the user’s history, preferences [1].

Typically in a recommender system, there is a set of users and a set of items. Each user u rates a set of items by some values. The chore of a recommender system is to predict the rating of user u on an un-rated item i or recommend some items for user u based on the existing ratings.

In rest of this paper, Section II describes the various types of filtering techniques used in RSs, Section III describes the various challenges faced by CF based RSs, Section IV describes the various approaches used with CF to overcome its challenges and Finally Section V concludes the paper.

II. TYPES OF FILTERING

There are various data Filtering techniques available for Recommendation Systems. This section describes these various techniques of Filtering. Recommender systems can be divided in four categories based on how user profile information is used: Demographic Information Filtering, Content Based Filtering, Collaborative Filtering and Hybrid Filtering.

A. Demographic Filtering

Demographic information means the personal attributes such as age, gender, livelihood, nationality which describes the individuals. This Demographic information may be used to classify the user on the basis of common personal attributes. Recommendation is carried out based on this Demographic Categorization [3]. Demographic filtering is relying on the principle that persons having common demographic information will also have common preferences [4]. The main advantage of this filtering is that, it may not have need of user rating history as it is totally based on Demographic information of user [3].

B. Content Based Filtering

Content based filtering is absolutely depends on the historical data of users choices. Content based approach...
recommends services analogous to services which are formerly preferred by user. The content based filtering methods based on information about the items or attributes of items which are available for recommendation. This set of attributes which typify the item within system is nothing but the Item Profile [5]. The system creates the content based user profile based on the weighted vector of item features. Weight denotes the importance of each feature to the user [1]. These weights are computed from the individually rated content vectors using a variety of techniques such as average value of ratings, Bayesian classifier. The Content based recommendation process basically consists in matching up the attributes of the user profile adjacent to the attributes of a content object i.e. item profile. The result is a relevance verdict that represents the user’s level of interest in that object or the likelihood that the user is going to like that object.

Content based filtering is straightforward and effectual but it also poses some negative aspect

- An adequate amount of ratings have to be collected before a content-based recommender system can really understand user preferences and provide precise recommendations. Therefore, when a small number of ratings are available, as for a new user, the system will not be able to provide unfailing recommendations [1].

- Content-based recommenders have no inherent method for finding users new interests [1]. The system put forward items whose scores are high when matched against the user profile.

C. Collaborative Filtering

Currently most Recommender System uses this Collaborative Filtering method as a recommender algorithm [6]. Collaborative filtering methods based on collecting and analyzing a outsized amount of information about user’s behavior, activities or preferences and predicting what user’s will like based on their likeness to other users [1]. The collaborative filtering method is based on the principle that if two users have same or almost same commonly rated items, then they may have similar preferences or tastes [7]. Such users are called as similar users. The basic aim of Collaborative filtering method is to build the user-item matrix by collecting user preferences or activities and try to find out the users with same interest. The users which have the similar interest forms the group called as neighborhood. The items which are unrated by the user but rated by his neighborhood are recommended to this user [7].

There are two main approaches used for Collaborative Filtering Techniques

a) Memory-based Approach:

Memory based collaborative filtering approach applies the whole user-item collection in dataset to make predictions [5, 6]. The main aim of this approach is to find out the similar user or services based on the previous rating stored in a user-item dataset. These similar users form the group called neighborhood. This neighborhood sets can be find out by various techniques such as Pearson Correlation Coefficient, Cosine Distance and Euclidean Distance [1, 5]. After finding out this neighborhood sets, the ratings of the user from this set are combines to make new predictions of unrated items to the user. The examples of Memory-based CF are User-based and Item-based CF [2]. Distinctively, user-based CF methods recognize users that are similar to the queried user, and guesstimate the desired rating to be the average ratings of these similar users. Similarly, item-based CF recognize items that are similar to the queried item and guesstimate the desired rating to be the average of the ratings of these similar items.

b) Model-based Approach

Model-based approach is based on the making of model from the on hand training dataset. This approach uses only the training dataset as an alternative of whole dataset to generate the model [1, 5, 6]. This model generated is then used to present predictions about the ratings of items which are unrated by user. Consider the example of Cluster-based Collaborative Filtering method. It builds the cluster of users as a model for further rating predictions. The ratings of users within the cluster are used to predict the ratings of unrated items by user. Other examples of model-based approach are Decision trees, Aspect Models and Latent Factor Models [1, 2].

D. Hybrid Filtering

For efficient and accurate results, numerous Recommendation Systems combines different filtering techniques with each other [7]. In this Hybrid method the combinations like Collaborative Filtering with Demographic Filtering, Collaborative Filtering with Content-based Filtering can be used. In another way we can encompass probabilistic methods such as Clustering, Decision tress into the Collaborative Filtering [6]. These combinations of different approaches proceed in a different manner to accomplish the desired goal.

III. CHALLENGES OF COLLABORATIVE FILTERING

The Collaborative Filtering is proven successful and widely used in the RSs. This Section mainly focuses on the challenges faced by the Collaborative Filtering approach.
A. Scalability

The number of users and products or services is ever-increasing enormously. The existing computational resources are scarce for processing this colossal information and forming recommendations. Majority of these available resources are consumed to determine similar user with similar preferences or items with similarity descriptions [7]. But the numerous systems have to react instantaneously with online requirements and bring up recommendations to all users [1]. These requirements cause the scalability problem when colossal amount of products or services and users are available.

B. Data Sparsity

Many of the E-commerce sites have the number of items and the number of users buying these items. Among these users, there are always numerous users who rated very few items [7]. Collaborative Filtering always uses this rating history of users to identify the users with similar predilection and create vicinity of the user. But user has very few items rated, then it’s becoming grueling to find users leaning and it may lead to user in different vicinity set [2, 7]. Thus this problem of scarce information of user rating is called as Data Sparsity problem.

C. Accuracy

Accuracy in recommendation system plays an eminent role. Accurate recommendation helps user to go for the products or services according to his requirements and helps the E-commerce systems to enhance the business [9]. The accuracy problem mainly occurred due to Data Sparsity [2]. Besides this, many recommendation systems use the user immaterial information for computation and this will also leads to imprecise recommendations.

D. Cold Start

The collaborative filtering mainly based on the users rating history and if this rating history is scarce it will lead to inaccurate recommendations. When any new user enters into the system, it is grueling to find similar users to him as scarce availability of information [1]. It is true in case of new items as well. New items cannot be recommended until it has some rating from users. Both of the above situations called as a Cold Start problem [7].

E. Privacy

Many people do not crave any intervene in their life. They do not want their habits, preferences, tastes and views to be open for others [8]. Thus Privacy is becomes a notable issue. In order to make the recommendation accurate and ideal for, it is important to acquire user data as much as possible [7]. If the Privacy about user data is ensured to the users, they will feel happy to make their information available to the others [2]. Thus assuring privacy will made RSs more effective.

F. Trust

If there is no restraint on who can give ratings, many companies made counterfeit ratings to make their product positive over other company products [8]. Sometimes CF uses this counterfeit rating and made recommendation to the user, which may not be ideal for users. So Trust on reviews is also an important factor for RSs.

IV. RELATED SURVEY

We study the various research papers and journal and come to know about various approaches used with CF techniques to surmount challenges mentioned in Section III. This section describes few of these approaches with the name of authors and their respective title.

A] By Xiwang Yang, Yang Guo and Yong Liu [9] “Bayesian Inference-Based Recommendation in Online Social Networks”. This paper focuses on the precise and personalized recommendation. The paper proposed a Bayesian-Inference-Based recommendation system based on social networks for precise and personalized recommendation. In the recommendation system, accurateness plays an imperative role. Accurate recommendation allows user to go for an ideal and desired item from the accessible information without inundated by immaterial information. Nowadays the social network site becomes a reliable network. The people are relying on the opinion of socially connected people for purchasing and consuming service or items. The opinions can be effortlessly shared by using this kind of social networks. Taking the benefit of this, the paper proposed a Bayesian-Inference-Based Recommendation algorithm. The algorithm operates on top of an underlying network and can be used for recommending general products or services. The algorithm is illustrated using the movie recommendation application. The people who are connected using these social networks share their ratings of movies among them. The proposed algorithms keep track of this shared rating history between friends. Based on this rating history, rating similarity between friends is carried out using Conditional Probability Distributions. According the calculated rating similarity, the Bayesian-Inference-Based frameworks carry out the plausible personalized recommendation for active user. The experimental results shows that the accurateness of Bayesian-Inference-Based recommendation is better than previous centralized Collaborative Filtering and Trust Based approaches.

Filtering Approach for Big Data Application”. The Collaborative Filtering (CF) techniques are very effectual and used by various e-commerce RSs, but CF faces few challenges such as recommendation within satisfactory time and idyllic recommendation with desired exactness to the users. In CF the similarity computation between every pair of services or users is a decisive and time consuming step. It step over the processing potential of RSs and resulting into lose of timeliness of RSs. While computing rating similarity, CF may consider many services immaterial to the user which may influence the exactness of predicted rating. To surmount above challenges, this paper mainly focuses on dropping the online execution time of CF and improving its accuracy. The Clustering-Based Collaborative Filtering approach proposed in this paper is focuses on gathering similar services into the single clusters and recommends services from cluster collaboratively. The methodology used in this paper is alienated into two main phases. In first phase, instead of direct applying collaborative CF to available services, services are recruited into some clusters based on their similarities using AHC algorithm. In second phase the CF is applied within a cluster to compute the rating similarity and recommend ideal services to the user. Since Clustering lessen the number of services available for recommendation, the time required for CF to compute rating similarity may be reduced significantly. As all the services within the clusters are pertinent to each other, the rating similarity between them is also more relevant and accurate. Hence accuracy of RSs may be enhanced.

C] By Shunmei Meng, Wanchun Dou, Xuyun Zhang and Jinjun Chen [11] “KASR: A Keyword-Aware Service Recommendation Method on MapReduce for Big Data Applications”. In the last few years there is a magnificent raise in the amount of products, customers and services. This enhancement leads into the Big Data analysis problem for the Recommendation Systems (RSs). The existing RSs are incompetent to process this massive amount of data and may undergo scalability problem while processing such a big data. The existing RSs recommend services or products to the diverse users based on same rating and ranking of services. It did not care about any specific user preferences and hence it is not much useful for personalized recommendation. To address the above challenges, this paper proposed a Keyword- Aware Service Recommendation (KASR) method. The main focus of the KASR is to provide Personalized Recommendations to the user depends upon each user preferences and hence to augment the accuracy and effectiveness of RSs. The different user preferences are specify by the keyword set. This keyword set are used by user based CF algorithm to generate ideal recommendation to the user. Keyword sets are generated by the help of Keyword-Candidate List and Domain Thesaurus. The preferences for the active user are given by him by selecting Keywords from Keyword-Candidate List and the preferences of previous user are collected by digging out their previous reviews. To deal with the scalability and efficiency issue, the KASR is implemented on a MapReduce framework i.e. Hadoop. The experimental results show that the accuracy and scalability of KASR is better than the existing approaches.

D] By Zibin Zheng, Hao Ma, Michael R. Lyu and Irwin King [12] “QoS-Aware Web Service Recommendation by Collaborative Filtering”.This paper mainly focuses on utilization of Quality-of-Service (QoS) values of web services experienced by different users. QoS values mean the non-functional characteristics of web services. As the popularity of web services in the World Wide Web is increasing day by day, the QoS value becomes more imperative to identify the most favorable web service from the available web services. On the basis of this, the paper propose a Collaborative Filtering(CF) approach for prediction of QoS values of Web services and recommend idyllic web services to the users by taking advantage of past usage experiences of web service users. In the previous approach the user makes the evaluation of web services by real time invocations of them, and observes the QoS values of web services. On the basis of this observed QoS values user go for the optimal service for him. But this previous approach is time consuming and costlier as many services may be charged for their invocations. Thus in previous approach, to obtain accurate QoS values without user invocation is grueling. Consequently the optimal web service selection and recommendation is becomes arduous. To attack these challenges, this paper proposed a CF based approach to predict the personalized QoS values for the web services. Firstly paper proposed the user collaborative mechanism to collect the historical QoS values of web services experienced by the different web service users. On the basis of this obtained QoS values, this papers proposed a method by combining user-based and item-based PCC approach to obtain a predicted QoS values of web services for the active user. The proposed approach in this paper is useful to obtain the QoS values of target services without real world invocations of it and also helps to recommend most favorable web services to the active user. The experimental result shows that the prediction accuracy of this approach is superior to other existing approaches.

E] By Hyung Jun Ahn [13] “A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem”.Collaborative Filtering (CF) technique is the most successful and extensively used by many Recommendation Systems (RSs). The foremost chore of this CF is finding similarities among users based on their rating history. This chore of similarity measurement is mainly carried out by Traditional distance and vector similarity measures such as Pearson Correlation and Cosine measures. But these Traditional measures have some limitations. These measures do not use any domain specific meaning of user’s rating
history. This leads to the Cold-start problem if scarce user rating data is available and consequently degrades the quality of recommendation. This paper mainly focuses on addressing these limitations of Traditional measures and consequently to overcome the Cold-start problem. Lot of studies is done by different authors to focus on the problem of Cold-start problem. But many of these studies put forth the Hybrid Recommendation approach using both Content-based and Collaborative techniques. Content-based technique makes use of the content information of items for new user if no user rating data is available. But the study carried out in this paper is somewhat different. This paper presents a new Similarity measures called PIP measure. PIP utilizes only the domain specific meaning of user’s rating. The advantage of PIP measure is that, it does not require any other information rather than user’s rating data unlike the previous Hybrid approach in which content information of item is used for recommendation. The experimental results show that PIP has better performance for users who creates the Cold-start problem.

F) By Weike Pan, EvanW. Xiang, Nathan N. Liu and Qiang Yang [14] “Transfer Learning in Collaborative Filtering for Sparsity Reduction”. Collaborative Filtering is mainly based on user-item matrix created from user rating history. It uses the existing values from user-item matrix and predicts the missing value from this user-item matrix. But in case of new user or user who does not access so much items, ratings to many items is not available from these users. This leads to the Sparsity of the user-item matrix. Sparsity of Data means the information available is either scattered or very few. This Data Sparsity creates a problem to CF and disgrace the quality and performance of Recommendation. In this paper the main focus is to address this Data Sparsity problem. Many related researches are carried out by different authors to surmount this Data Sparsity problem. These researches put forth the Transfer Learning approach to unravel this aforementioned problem. This Transfer Learning method is based on the utilization of the data from the other available and Recommendation Systems called as Auxiliary domain. Transfer Learning method transfer the related and useful data from auxiliary domain to the target domain i.e. current RS. But this previous method has some confines such as the user rating from both the auxiliary ad target domain should be identical and both user and item from auxiliary domain should be related to the target domain. To addresses the aforementioned limitation, this paper propose a matrix factorization based framework called Coordinate System Transfer (CST). This CST is useful for transferring both user and item knowledge from an auxiliary domain to target domain and hence to overcome the Data Sparsity problem. The basic purpose of CST is to discover the common latent information from auxiliary and target domains. The CST is based on the postulation that there always exists a finite set of tastes which typify the domain independent rating structure of users called as Principle Coordinate. This Principle Coordinate is used to define Common Coordinate System for representing users. The CST by utilizing the Common Coordinate System and Sparse Matrix Tri-factorization finds the proper auxiliary domain for knowledge transfer. Then CST utilizes the Regularization technique to adopt the transferred knowledge into target domain. The experimental result shows that, the CST has superior performance than other present methods which developed for addressing the data Sparsity problem.

V. CONCLUSION

In this paper, we first introduce the RSs and the filtering types used for RSs. From various types of filtering techniques, the CF is extensively used by numerous RSs. But this CF technique is facing many challenges. We also introduced many of these challenges in this paper. Further in this paper, we attempt to acquaint with a comprehensive survey of various approaches used in combination with CF to addresses the mentioned challenges. There have been many researches and authors proposed the solution with results which concentrates on individual challenge. Despite of all these researches, current RSs requires some more improvements to make recommendation more effective for wider range of applications. This further required improvement can be achieved by combining few of these aforementioned researches in one approach. It is expected that many of these mentioned CF challenges will be overcome by this single, combined approach.

REFERENCES


