Distributed Multi Agent Driven Collaborative Filtering Based Recommender Systems

B.Raghuram*, S.V.Krishnarao*, G.Srider*

#Department of Computer Science &Eng
Kakatiya Institute of Science and Technology
Warangal, India.
raghu9b.naik@gmail.com
Svkrishna_rao@yahoo.com
sridhar_gujjeti@yahoo.com

Abstract— The huge spread of Internet technologies in today’s world has imposed a great challenge and burden on users to utilize information properly. This made the recommender systems popular, which provide opportunities to examine and collect information about various things even without exploring all of them individually. Collaborative Filtering (CF) is one of the mostly used techniques in the field of recommender systems which provides personalized recommendations to users based on their previously expressed preferences and those of other similar users. The great drawback of collaborative filtering is that it is unable to exploit multiple distributed information resources. This information may be different domains with either heterogeneous or homogeneous repositories. To overcome this drawback, we used a distributed multi-agent communication infrastructure that can mine the information from distributed heterogeneous repositories to facilitate base to run collaborative filtering algorithms. Along with this our model is not specific for any single type of collaborative filtering algorithm instead the model is capable of running any type of collaborative filtering algorithms in coordination with any data mining algorithms. Experiments conducted on Java Agent Development Environment (JADE) using movie goer’s data sets demonstrate that our approach is practically consistent.

Keywords- Recommender Systems, Collaborative Filtering, Agent Technology, Distributed data minin.

I. INTRODUCTION

Collaborative Filtering [1] is recommender system technique used in many E-Commerce recommendations to support users for the selection of different commodities which are favourable to them and other similar user [2]. Collaborative Filtering works with an assumption that people with similar tastes also prefer the same kind of items. To generate a recommendation with respect to a user, the collaborative filtering model creates a neighbourhood of users with the highest similarity to the user whose preferences are to be predicted. It then generates recommendations by calculating the normalized and weighted ratings of the users previous data and using the neighbourhoods data. In this process make use different data mining techniques also.

In today’s world the, information about the users and their exploration is vastly distributed among many data repositories, in a variety of domains. When integrated, these repositories could provide a recommendation, while a CF based on a single repository may fail to do so. In this work we discuss the details of operating CF over a distributed setting of data repositories and compare different distribution approaches. We propose the architecture of the above ideas using distributed multi agents which can provide transparency to users in case of heterogeneity. Multi agent driven system supports group communication in multi-agent systems, where similar service-providing agents are connected into a single virtual structure called “implicit organization” [3]. It is all ruddy proved that the agents can perform collaborative filtering task if multiple agents engaged for different tasks [4].

To make our proposal more realistic and applicable to real world domain we conducted several studies. In concern to that we studied the impact of different data distribution scenarios on the quality of recommendations. We studied two types of distribution representing in real time application of web usage, they are general and relevant distributions. General distribution - imitates a situation where information about a most of user is available to close and other vicinity. In this scenario, each agent organization and collect data form heterogeneous sources. Relevant distribution imitates a situation where each repository stores information related to a limited number of topics.

The data mining techniques very frequently getting used for effective results in collaborative filtering models. But in case of distributed environment using the data mining techniques is not easily possible because the data in distributed environment might stored in heterogeneous data bases and the format of the data may also differ from one source to another source. This emphasis the requirement of agent which can take care of heterogeneous sources. Using of agents in collaborative filtering based recommender system already exists but many of models are specific to particular type of algorithm which may not provide the environments where different models of collaborative filtering and data mining.

From the literature review we found that even though lot of research progressed in the field of collaborative filtering there is a very few of proposal concentrating on collaborative filtering on distributed sources and even this proposals not provide efficient solution for heterogeneity of data and most proposals are specific to a particular type to overcome such drawbacks we proposed a distributed multi agent based model which can handle heterogeneous sources and provide platform
independent architecture which can work with any type of collaborative filtering and data mining models.

The rest of the paper is organized as follows. Section 2 reviews the related works on distributed Collaborative Filtering and their applications. In Section 3 we discuss about distributed agents and their collaborative filtering advantageous. Section 4 gives our distributed multi agent collaborative filtering architecture and process. Section 5 presents experimental results. Finally, we conclude and present the directions of future research.

II. RELATED WORK

All implementing the collaborative filtering algorithms in a decentralized way was initially proposed in [5]. It presents a distributed architecture supporting product recommendations for different customers represented by software agents. But the problem with this proposal is the communication between the deployed agents used an expensive routing mechanism based on network flooding that increased the communication overhead. A matured mechanism was proposed in [6] even it reduced the efficiency of the neighbourhood formation phase. The proposal in [7] elaborated on the discussion of distributed collaborative filtering models. It developed a detailed taxonomy of distributed Collaborative filtering in recommender systems and presented different implementation frameworks for different domains of E-Commerce and E-Business system. Most of these studies did not include thorough experimentation and did not analyse the different factors that might affect the quality of the generated recommendations.

The recent empirical research shows that due to the limited information-processing capability of the human mind, users tend to rely heavily upon the personalized agent’s product recommendation to make a purchase decision [8]. Therefore, it is crucial for e-commerce/business/service sites to provide knowledgeable recommendation agents to online users. From the above study we conclude that even though it is proved that collecting information for collaborative filtering from various sources provides efficient result but there is lack proposal which can provide solution to problem of heterogeneity the cooperative multi agents can be used to get efficient results.

III. DISTRIBUTE AGENTS & COLLABORATIVE FILTERING

Even though the Collaborative filtering is a most familiar and most widely implemented recommendation technique it relies on the idea that people who agreed in the past will also agree in the future [9]. The input for the collaborative filtering recommender system is a collection of user ratings for items, where each row represents the ratings of a single user and each column represents the ratings for a single item. Collaborative filtering aggregates ratings of items to recognize similarities between users, and generates a new recommendation of an item by weighting the ratings of similar users for the same item [10]. The main advantage of CF is that it is completely independent of any item representation. Thus items can be recommended regardless of their contents and same makes it applicable for distributed environment.

Agent technology is one of the most promising technologies for facilitating personalized recommendations. Software agents are being used in an increasingly wide variety of software applications ranging from comparatively small systems such as personalized e-mail filters to large complex mission critical systems such as air-traffic control [11]. The notable characters associated with software agents such as autonomous, pro-active, goal-oriented, intelligent, social, etc. make software agents well suited for playing the role of personalized recommenders to individual users of various e-service, e-commerce and e-business sites. The cooperative agents which can collaborate with each other to perform a particular task useful in distributed application. In case of distributed applications the same type of task may need to carry out at various sources in coordination in such cases the cooperative agent play vital role.

In case of collaborative filtering to gather subjective data from various sources and to handle heterogeneity the cooperative multi agents can be used to get efficient results.

IV. DISTRIBUTE MULTI AGENTS DRIVEN COLLABORATIVE FILTERING ARCHITECTURE

Analytical Mining Controller Agent: This is the core agent of the distributed knowledge discovery system. The function of this agent is to dispatches the mobile agent to each sub-dam and obtains the frequent item set information from local data.
base and storing the local frequent item sets information in to data warehouse to get the global knowledge.

Centralized Store: The history of collected information from different sources is stored in the centralized store. The databases in the different sites of the distributed information system are often heterogeneous. The data centralized store should transfer the data to uniform format so the data can be organized and accessed efficiently

Collaborative Filtering Engine: this engine capable of running collaborative filtering algorithms on user profiles available in centralized store in coordinating with DM engine.

Data Mining Engine: Data mining engine capable of running different mining algorithms in coordinating with CF engine.

Collaborative Filtering DB: The collaborative filtering database will store the result to present to users which are results of CF engine and DM engine.

Data Cube: The data cube provides the feature to analyse the centralized data with various levels of generalizations.

Mobile Agent Controller: MAC is the framework within which the mobile agent activities in the distributed data mining system take place. MAC is responsible for generating, activating and assembling the agents required for the data mining process. The different agent types and their tasks are briefly discussed below.

Global Information Extraction Agent (GIEA): This agent can be dispatched to centralized store to extract the global data which is requiring by data cube.

Data Subjectivity Check Agent (DSCA): This agent can be dispatched to sub-DAMS to select subject oriented data with respect to data centralized store out of the various data fields available in local databases.

Data Consistency Check Agent (DCCA): This agent can be dispatched to sub-DAMS to checks whether data retrieved from local data bases are in compatible format with centralized store format or not and pre-process the data in the local database to the format of data warehouse.

Data Reliability Check Agent (DRCA): This agent can be dispatched to sub-DAMS to checks whether local databases reliable source or from unreliable source and intimate the information to data.

B. The Multi Agent System Organization

In the proposed system to provide recommendation for user the data from general(open to all type of users) and relevant data source collected by agents. In order to handle heterogeneity different agent will work with collaboration on the commands of Mobile controller agent. The data reliability check agent decide the reliability of data source and decide whether to collect or not in case of positive result it sent consistency check agent which handle different platforms and convert the data as per require format and send to agent which collects the subjective mater from the source and it sends it to global information extraction agent which will be sent to centralized store.

Using data from centralized source the data cube can be formed or directly by using the data from the centralized source the collaborative filtering algorithms can be applied. In case of the using data mining algorithm it can be performed by collaboration of Collaborative Filtering-engine and Data Mining-engine. In this model both collaborative filtering engine and data mining engines are independent of specific model. So any algorithms of collaborative filtering or data mining can be used in this proposed architecture.

C. Generating Recommendations

First of all to generate recommendation the system will predicted scores of various recommendable items. After producing the scores the recommendation process proceeds to determine which of the items are actually recommended to the particular user. To do this our model uses mining algorithms on collected data from various sources. To collect the data from various heterogeneous and distributed sources the above mentioned agents will be used. As explained different agents will process the data in different formats and information will stored in centralized source. There the collaborative filtering engine will perform the collaborative filtering using the data mining engine and give final recommendations for users.

D. Recommendations using Multiple-level Similarity

Multi level similarity is utilized to produce recommendations when the known preferences of the user are very limited and, consequently, the number of recommendable items may be very small. This follows following steps:

1. Find the preferred categories of the active user from various available sources.
2. Determine association rules from such categories.
3. Find rules containing items in those associated categories and assign predicted scores to items in their heads.

The above mentioned three steps of collaborative recommendation system will make use of various data mining algorithms as per requirement. We designed system as such it can apply using any data mining techniques. Finding the preferred categories of active user is required to engage the classification and clustering algorithms in order to identify different user categories and different users belongs to those categories. Out find categories and active users involve in that it is required to find set of common associations among them this process usually requires applying different kinds of association rule mining techniques. Using this common association rules the recommendations will be generated.

Finally the using association recommendations of different categories the similarity scores need to be assigned to different predicted scores are weighted by the User's preference on its categories. To predict scores of a active user the rule associated recommended generated will cross checked against the users previous history. If the users previous history matches with the recommended associations such rules will be given highest predicted score if the rule matches with less similarity it will be given less preferred rules. The recommendation with highest predicted score will be given high preference. The predicted preference of an item can belongs to more than one category in such cases the average of the preferences will be considered for all categories.
To perform preferred match among generated rules and users' profile and to generate predicted score any popular collaborative filtering model can be used. The final recommendations will be displayed to user based on threshold value set by user or it can be determine by system itself based on user's previous profile. The recommendations with predicted scores more than threshold will be displayed as recommendations to user.

V. PERFORMANCE EVALUATION

The experiments conducted using P4-2.0 Ghz processor and 512 MB RAM based PC's in Java Agent Development Environment (JADE) and we used data movie recommended data sources from reference [12]. To provide heterogeneous environment the collected data is stored on different PC's with different type of data bases. There are several commonly used evaluation metrics for evaluating CF algorithms. They include accuracy measure precision and recall, and the percentage of recommendable items in the system known as coverage. The choice of these performance metrics is dependent upon the recommendation strategies used.

Experiment conducted in a fold cross validation, and the results of all trials were averaged to obtain the result. In each trial, one preference data of each user in the test set was hidden. For each active user, a list of Top-N recommendations was generated based on the user's known preferences. If the hidden item was recommended, it is called a hit. The recall rate of an algorithm is deepened as the number of hits over the number of hidden items in the test sets.

The various measurements that can be used to predict the scores of recommendable items. These measurements include the product of support and confidence. The results of an empirical comparison between these measurements. The confidence shows 91.7%, 70.2% and 51.3% for all values of N respectively on different movie recommended datasets [12]. This shows that the correlation of items is not as important as the strength and interestingness of rules when producing collaborative recommendations.

VI. CONCLUSION

Considering the lack of proposals that can provide collaborative filtering recommendations from heterogeneous distributed sources and can work with any type of collaborative filtering and data mining algorithms we proposed a agent based architecture that can handle heterogeneous distributed sources and can work with different collaborative and mining algorithms. The experiment conducted on movie data base using JADE shows results to our model is effective. Developing the models for recommended systems with more complicated collaborative filtering algorithms on proposed model will be our feature perspective of research.

REFERENCES