International Journal of Computer Applications Technology and Research
Volume 2– Issue 6, 650 - 659, 2013, ISSN: 2319–8656

Reviewing Cluster Based Collaborative Filtering Approaches

F.Darvishi-mirshekarlou
Department of computer, Shabestar branch, Islamic Azad University
Shabestar, Iran

SH.Akbarpour
Department of computer, Shabestar branch, Islamic Azad University
Shabestar, Iran

M.Feizi-Derakhshi
Department of computer engineering, Tabriz University
Tabriz, Iran

Abstract: With regard to rapid development of Internet technology and the increasing volume of data and information, the need for systems that can guide users toward their desired items and services may be felt more than ever. Recommender systems, as one of these systems, are one of information filtering systems predicting the items that may be more interesting for user within a large set of items on the basis of user's interests. Collaborative filtering, as one of the most successful techniques in recommender systems, offers some suggestions to users on the basis of similarities in behavioral and functional patterns of users showing similar preferences and behavioral patterns with current user. Since collaborative filtering recommendations are based on similarity of users or items, all data should be compared with each other in order to calculate this similarity. Due to large amount of data in dataset, too much time is required for this calculation, and in these systems, scalability problem is observed. Therefore, in order to calculate the similarities between data easier and quicker and also to improve the scalability of dataset, it is better to group data, and each data should be compared with data in the same group. Clustering technique, as a model based method, is a promising way to improve the scalability of collaborative filtering by reducing the quest for the neighborhoods between clusters instead of using whole dataset. It recommends better and accurate recommendations to users. In this paper, by reviewing some recent approaches in which clustering has been used and applied to improve scalability , the effects of various kinds of clustering algorithms (partitional clustering such as hard and fuzzy, evolutionary based clustering such as genetic, memetic , ant colony and also hybrid methods) on increasing the quality and accuracy of recommendations have been examined.

Keywords: Information overload, Recommender systems, Collaborative filtering, Clustering, Fuzzy clustering, Evolutionary based clustering

1. INTRODUCTION

Due to expansive growth of the World Wide Web as well as increasing the amount of available information for each person, some problems have appeared for users in identifying useful, required and interesting information for each person. In most cases, people are faced with choices and very large data volumes, and searching all of them is out of user' capability. This problem is called information overload [1]. With regard to increasing high volume of data on the Internet, users have encountered with the problem of finding the right product at the right time. Finding final data on the basis of users' needs has became complicated and a time consuming process. In response to this growing epidemic, especially e-commerce, recommendation systems have been proposed .These systems are personalized technology for filtering information. In daily life, often some suggestions and comments of friends are used in choosing something. In addition, use our previous experience is used in selecting a specific item. Sometimes, it has happened that, on the basis of the suggestion of friends, we have bought a book or a particular product, or we have watched a film or have listened to music.

There are lots of e-commerce businesses in which one or more variations of recommender system technology is utilized in their web sites like Amazon.com (book recommendation system), movielens (move recommendation system), ringo (music recommendation) and etc. These systems use knowledge of the user's interests. This knowledge is obtained from searching the web pages for finding their favorite items or pages.Recommender systems are classified into three fundamental groups, namely, content based model, collaborative filtering and hybrid systems. Content based systems save content Information and product specifications, and then provide some suggestions for users on the basis of the similarity between customer's purchase history and other items. Collaborative filtering (CF) act is based on the experience obtained from the purchase history of similar customers [4]. Hybrid systems try to combine above models in different ways to overcome the problems that have appeared because of using both content and collaborative filtering. Also these systems improve recommendation performance [26, 27, 28]. A brief review of collaborative filtering and its challenges are mentioned in section 2 and some of recent collaborative filtering approaches that are based on clustering are reviewed in section 3. In section 4 we give the conclusion of this work.

2. COLLABORATIVE FILTERING

Collaborative filtering, as one of the most successful techniques, is based on the assumption that people who have similar interests in terms of some items, they will have the same preferences in other items. So the goal of collaborative filtering is to find the users who have similar ideas and preferences or to find the nearest neighbor of them. This method is carried out in three steps: preprocessing, similarity computation and prediction / recommendation generation.

In preprocessing step, user-item matrix is built. This matrix contains the ratings that represent the expression of user's preferences. These preferences are explicitly obtained by rating the product in (1-5) scales, or implicitly by their purchase history [3].

Table1. A sample of user-item rating matrix

rating	Item1	Item 2	Item 3	Item 4
User 1	5	5	1	1
User 2	4	5	1	2
User 3	1	1	5	5
User 4	2	1	5	4
User 5	1	1	1	3

In similarity computation step, statistical techniques are used to find similar users with active user on the basis of their similar past behaviors. It reflects distance, correlation, or weight between two users. There are many different similarity measures to compute similarity or weight between users or items such as Cosine Vector, Pearson Correlation, Spearman Rank Correlation, Adjusted Cosine and etc. Pearson correlation, as the most common measure, has been mentioned.

$$W_{u,v} = \frac{\sum_{i\in I} (r_{u,i}-\bar{r}_u)(r_{v,i}-\bar{r}_v)}{\sqrt{\sum_{i\in I} (r_{u,i}-\bar{r}_u)^2}\sqrt{(r_{v,i}-\bar{r}_v)^2}} \quad (1)$$

$W_{u,v}$ is the similarity between two users. u and v. I is the set of co-rated items by tow users. $r_{u,i}$ and $r_{v,i}$ stand for the rating that has presented by user u and v in item i. $\bar{r}_u$ and $\bar{r}_v$ refer to the average rating of the co-rated items of the u th and v th users respectively.

In prediction step, weighted aggregate of similar user's ratings are used to generate predictions for active user. Finally, after predicting rating for items that have not been observed by active user, recommendation has been generated, and a list of items with high rating has been recommended to user.

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u\in U} (r_{u,i}-\bar{r}_u) . w_{a,u}}{\sum_{u\in U} |.w_{a,u}|} \quad (2)$$

$P_{a,i}$ is the predicted rating for the active user, a, on a certain item, i. $\bar{r}_a$ and $\bar{r}_u$ stand for average ratings for the user a and user u in all other rated items, and $w_{a,u}$ is the weight between the user a and user u.

2.1 Memory/model based CF

Collaborative filtering is grouped into two general classes, namely, neighborhood-based (memory-based) and model-based methods. In Memory-based CF systems, the whole user-item rating dataset is used to make predictions. This system can be performed in two way known user-based and item-based recommendations. User-based collaborative filtering predicts an active user's rating in an item, based on rating information from similar user profiles, while item-based method look at rating given to similar items.[2]

Model-Based system contains building models on the basis of extracting information from datasets that can predict without using whole dataset. In these systems, the complete dataset is merged into train and test dataset. Train set is utilized to train the model using various algorithms like Bayesian network, clustering algorithm, regressions, matrix factorizations etc. Then, the trained model is used to generate recommendations for active user in the test set. The test set must be different and independent from the training set in order to obtain a reliable estimation of true error. Management of low density of data set is one of the most important advantages of model-based system that improves the scalability of big data sets. Because of the off-line building of models, the response speed time will be decreased, and less memory is used. The high cost of building these models is disadvantage of model-based systems [2].

2.2 Challenges of CF

There are three fundamental challenges for collaborative filtering recommender systems such as Data sparsity, Cold Start and Scalability.

Data sparsity: this issue is take place when the user-item matrix is extremely sparse, that is, users rate only a small number of items, so accuracy of recommendation will be decreased.

Scalability: with development of e-commerce and growing the number of users and items in such systems, the Scalability will increase, and ultimately the prediction calculations will be prolonged. Dimensionality reduction, clustering and item-based collaborative filtering are more common ways to alleviate this challenge.

Cold-start: when a new user or new item enters the system, it is difficult to find similar ones, because there is not enough information about his/her history in system. To overcome this issue, the hybrid system is used commonly. In this system, both rating and content information are used for users or items for prediction and recommendation.

2.3 Evaluation metrics

Several metrics have been proposed in order to evaluate the performance of the various models employed by recommender systems. Statistical Accuracy metrics and Decision support metrics are two major evaluation metrics. Statistical Accuracy metrics measure how much the predicted rate is close to the true rating that is expressed by user. Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) are the most common error metrics [6].

In Decision support metrics, the accuracy of decisions in recommender system is measured. Recalling and precision are most common metrics obtained by classifying the recommendation according to table 2.

Table2. Recommendation outputs classification

Correct / obtained	Recommend	Not recommend
Recommend	True-Positive(TP)	False-Positive(FP)
Not recommend	False-Negative(FN)	True-Negative(TN)

$$\text{Precision}=\frac{tp}{tp+fp} \quad (3)$$

$$\text{Recall}=\frac{tp}{tp+fn} \quad (4)$$

$$\text{Accuracy}=\frac{tp+tn}{tp+tn+fn+fp} \quad (5)$$

Precision is the fraction of the recommended items that are interesting to users, and recall is the fraction of the items having recommended higher ratings [6].

2.4 Datasets

Most often used and freely available datasets for collaborative filtering are EachMovie (EM), MovieLens (ML) and JesterJoke datasets.

EachMovie is a movie rating data set collected by the Compaq Systems Research Center over an 18 month period beginning in 1997. The base data set contains 72916 users, 1628 movies and 2811983 ratings. Ratings are on a scale from 1to 6. **MovieLens** is also a movie rating data set. It was collected through the ongoing MovieLens project, and is distributed by GroupLens Research at the University of Minnesota. MovieLens contains 6040 users, 3900 movies, and 1000209 ratings collected from users who joined the MovieLens recommendation service in 2000. Ratings are on a scale from 1 to 5**. Jester Joke** data set is collected by Goldberg et al. The data set is much smaller than above datasets that containing 70000 users, but only 100 jokes.

3. CLUSTER BASED CF APPROACHES

Clustering, as one of the common techniques to grouping the object, is a promising way to improve the scalability of collaborative filtering approaches by reducing the search for the neighborhoods in the preference space, and generates some recommendations for users without using whole dataset. It means that after Clustering, users similar to active users are chosen from his own cluster instead of choosing them among all users. Fig1 shows the impact of clustering on reducing user-item matrix dimensions. By applying similarity calculation against clustering, time complexity reduces from O(N) to O(k) where k is number of clusters.

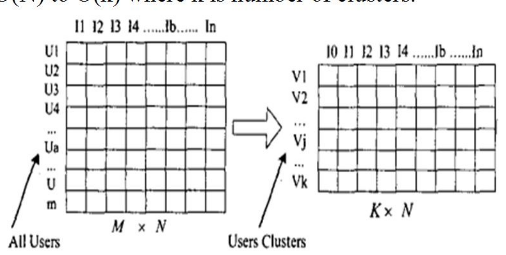

Figure 1. user clustering for collaborative filtering

There are many collaborative filtering based on partition clustering (hard/fuzzy) and evolutionary based clustering. In this paper, some recent approaches have been reviewed.

3.1 CF based on partition clustering

Clustering algorithms such as K-means assigns each user to a unique cluster. It is believed that this is too restrictive for most of the real world scenarios. In reality, users often have more diverse preferences. For example, one user may prefer both 'action movie' and 'comedy movie', or even more. By this consideration, it is more reasonable to allow a user to be assigned to more than one group. For this reason, fuzzy clustering algorithms are applied. so that users are clustered into different groups, and each user may be located in more than one group.

3.1.1 CF based on Fuzzy Clustering

An approach is proposed to improve Item based method in [7] employing FCM algorithm in item based collaborative filtering. In this method, items are partitioned into several clusters, and prediction is accomplished against clusters. By using clustering technique in Item-based, partitioning the data will be even less, so computational cost decreases. This approach shows that it overcomes scalability, cold start and sparsity. It saves more than 99% computational time, and does not change the prediction quality and eventually real-time prediction.

A new fuzzy algorithm is proposed in [8] employing Entropy based FCM in item based collaborative filtering. Despite of well accomplished IFCM [7], prediction results obtained from all clusters may include too much noise, and the accuracy would be affected. In Entropy based IFCM, higher value of degree of membership in objective function will exponentially increase, and it is distinct from lower value. Due to ability of using a wider range of fuzzifier values in IFCME, IFCME is more flexible than IFCM. The results show that IFCME improve MAE by 3.2% and 13.4% in single rate items in comparison to IFCM. Another fuzzy algorithm is presented in [9] who has formulated objective function with Exponential equation (XFCM) in order to increase the capability of assigning degree of membership. By the way, noise filtering is incorporated in XFCM, and noisy data are clustered differently in comparison to other Fuzzy Clustering. Thus, the centroid is strong in the noisy environment. The experiments show that centroid produced by XFCM becomes robust through improvement of prediction accuracy 6.12% over (FCM) and 9.14% over Entropy based FCM (FCME). Although using fuzzy clustering is more convenient to allow data to locate in multiple clusters [7, 8], this is not enough to make accurate recommendations because irrelevant data could be assigned to the clusters and overwhelm the rating predictions. In general, the ratings should be computed by using only ratings from relevant items. To overcome this issue, a new clustering algorithm is offered in [10] who reformulates the clustering's objective function with an exponential equation. This equation locates data to clusters by aggressively excluding irrelevant data. In this method, data that are farther from cluster, negative degree of membership is allocated. The negative membership degree indicates very low correlation of data and cluster. On the other hand, if the membership degree goes beyond 1, data truly belongs to cluster. Thus, these properties are used to filter out irrelevant data when the degree of membership is negative. The MAE results show that XFCM outperforms FCM by 5.2～9.8%, FCME by 1.0～6.1% and the item-based method by 2.7～6.9%.

An improved FCM algorithm is presented in [11] who strengths item clustering by injecting FP-Tree approach to it. The reason of using FP-Tree approach is to calculate means of nominal and set type data. This means that the E-Commerce data always have many kinds of data types, like numerical, nominal, set and etc. For example, a product contains a number of features, brand, main function and price that are respectively nominal, set and numerical. Traditional FCM algorithms are not able to handle these types of data. Therefore, in order to overcome this problem and handle these types of data, FP-Tree is used. In this way, in the repeated part of FCM, FP-Tree algorithm is used to category data feature. Then, average of numerical feature is calculated, and average is considered as mean. Finally, the mean of one cluster is obtained. After clustering items, cluster-based smoothing technique [24] is employed to estimate probability of unseen term and fill the missing values in data set. The goal of using smoothing method is to handle the sparsity problem in collaborative filtering. At the end, neighbors of active user are selected and prediction for an item is calculated. Experiment results show that this framework works well, and can efficiently improve the accuracy of prediction.

A framework is proposed in [12] to extend traditional CF algorithms by clustering items and user simultaneously. This is approximately like Co-Clustering (COC) problem [25] in which each user or item can only belong to a single cluster, but the main difference is that each user or item can be located in multiple clusters (MCOC). For example, in a movie web site, a user may be interested in multiple topics of movies, and a movie can be interesting for different groups of users from different aspects. So, FCM algorithm is used to obtain this goal. To combine subgroups with existing collaborative filtering in this approach, collaborative filtering algorithm is used in each subgroup, and then prediction results of each subgroup are unified. Finally Top-N recommendation is performed. Experimental results show that, when more subgroups are used, each subgroup becomes more density. Hence, data sparsity problem reduces in some CF methods. Also, short runtime of MCOC demonstrates its good efficiency. Using subgroups is a promising way to further improve the top-N recommendation performance for many popular CF methods.

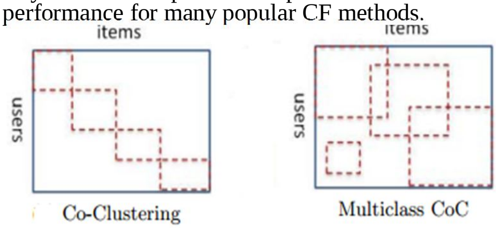

Figure 2. Comparison of COC and MCOC

A novel approach is suggested in [13] to alleviate the sparsity and scalability challenges in collaborative filtering recommendation. This approach firstly converts the user-item ratings matrix to user-class matrix. The user-classification matrix represents frequency expression of user preferences in terms of classification of items. Since the number of classifications is far less than that of the items, increase density of the data increases considerably in the resulted matrix. After converting matrix, FCM partitioning algorithm is applied to divide users into clusters. Finally, a list of Top-N recommendations is presented for each user. Computational experiments show that this approach is more effective and efficient than other well-known CF approaches, such as user-based CF and K- means-based CF.

New collaborative filtering method is proposed in [14] taking into account the impact of time on calculation of the users' similarity. It means that similarity of users is calculated in the same time or in similar periods. In this way, weight of time is assigned for each rating, and the weights of the most recent ratings demonstrate the latest ratings given by the users, and reflect the current interests of them, so the nearest neighbor will be found accurately. After finding similarity, FCM algorithm is used to clusters users. After clustering the user, fuzzy cluster-item rating matrix is constructed. It shows the rating given by a user cluster to an item. After calculating the dense user cluster-item rating matrix, similarities of items are calculated by choosing similarity measure. Finally, in recommendation generation, top N items that have the highest ratings are recommended to the user. Results show that the dimensions of the sparse user-item rating matrix are reduced, and improved algorithm can effectively raise the accuracy of recommendations.

3.1.2 CF based on Hard Clustering

A method is offered in [15] using cluster ensemble for collaborative filtering recommendation. In cluster ensemble result of some different clustering algorithms or result of several independent runs of one clustering algorithm are combined.

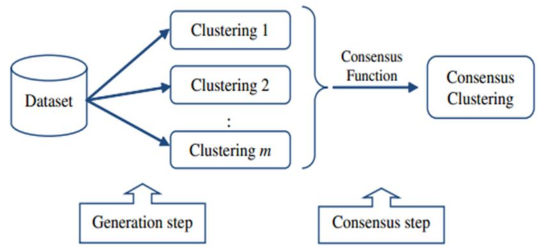

Figure 3. Clustering ensemble process

The main target in cluster ensemble is to have better performance than the single clustering algorithms. The goal of this approach is to show that employing cluster ensemble, in comparison to using single clustering, can improve recommendation accurately. In this approach, two base clustering algorithm, k-mean and SOM (Self Organized Map), are used. In this way, each clustering algorithm is run several times, and then the result of these runs is individually combined in census function. After clustering step, recommendation generation is performed once with SOM ensemble as well as k-mean ensemble individually. Finally, their results are compared with single k-mean and single SOM. Experiment results show that, by using cluster ensemble, recommendation accuracy and precise is high.

A hybrid method is proposed in [16, 17] over-coming cold-start problem in user and item re-spectively. The proposed method combines clus-tering and decision tree in which both the rating and content information are used. In this way and in the first step [16], items are clustered by k-mean clustering on the basis of ratings. K cluster obtained from clustering contains the items that are interesting for users with same preferences. One of the important content attributes of items in recommender systems is movie genres, such as action, comedy, romance, and etc. In the second step, decision tree is constructed in order to asso-ciate the new items with existing items. In this decision tree, genre information is used as attributes of tree, and clustering number obtained in the first step are used as the result to build decision tree. After building decision tree, in the third step, new item is classified. When new item without rating is entered to the system, the algorithm gains its attribute, and the item enters to decision tree. By following down the tree and answering the question correctly, the cluster number of new item will be achieved. Finally, in the last step, rating prediction is calculated, and recommendation is generated on the basis of this assumption that new item will be preferred by users who prefer the items in the cluster in which the item has been placed in the classifying procedure.

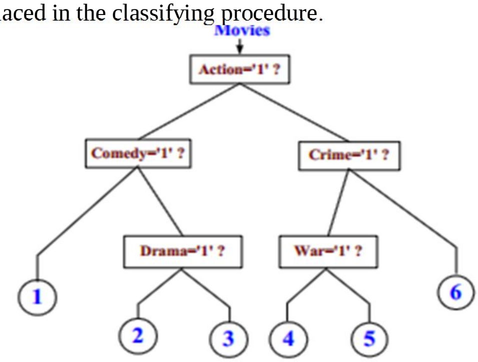

Figure 4. Example of decision tree in item side

According to [17], the proposed algorithm is used by user. Clusters are firstly used on the basis of ratings, and after that, decision tree is built by user demographic information (such as gender, age, occupation) and clustering numbers. The rest of algorithm is similar to [16], but it is used by user. Experiment results show that prediction accuracy is quite high in item and user's cold-start condition.

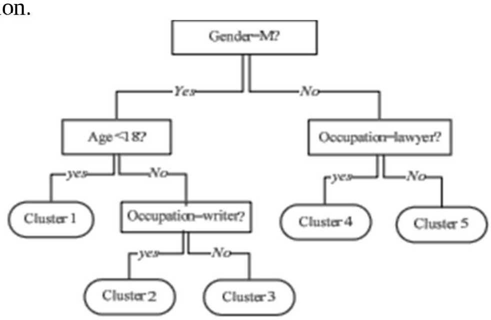

Figure 5. Example of decision tree in user side

A method has been proposed by [18] to alleviate sparsity problem in collaborative filtering in which item clustering and user clustering are used to generate recommendation simultaneously. This approach is performed in three phases. In the first phase, item based collaborative filtering is used for rating smoothing in which a full user-item matrix is obtained without non-ratings, and sparsity issue is resolved.

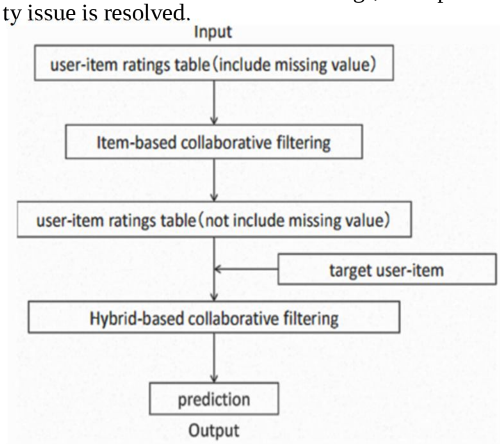

Figure 6. Stream of approach method

After rating smoothing, in the second phase, hybrid clustering technology is used to predict rating for target user. In hybrid clustering, users and items are firstly clustered in full user-item matrix. Prediction is calculated in the last phase. At first, users are found in the target user's neighborhood by user clustering result, and likewise items are found by item clustering result.

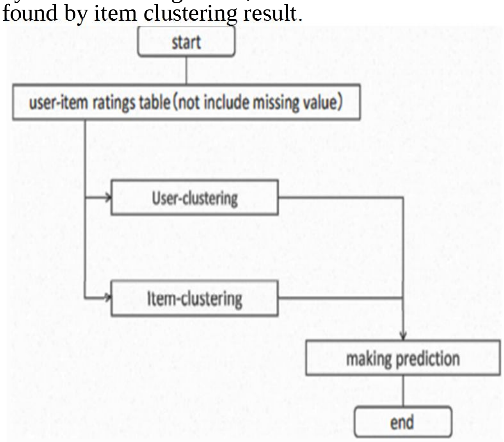

Figure 7. Hybrid Clustering Algorithm

Secondly, the user and item ratings are obtained from the full matrix. Finally, through those ratings, the item is recommended to target user. According to the experiment, it is concluded that the proposed collaborative filtering provides accurate prediction, and recommendation is more appropriate for the users because, in this method, the target user's neighborhood and target item's neighborhood are used simultaneously for recommendation.

3.1.3 CF based on evolutionary clustering

On of drawback that classical clustering algorithms are faced with it is falling in local optima, and in this case, evolutionary clustering algorithms are recommended to resolve it.

A hybrid clustering has been used by [19] for recommendations in which k-mean clustering and genetic clustering have been combined. The most important pivot point of the k-mean clustering is picking up the right initial seeds. An unsuitable choice of clusters may outcome poor results, so quality of clusters is dependent on initial cluster centers. In this approach, Genetic algorithm is used to pick up suitable initial seeds for k-means clustering. In fact, the objective of this approach is to pick up optimal initial seeds to produce high quality recommendations. After clustering, movies are recommended to target user which are interesting for users in the cluster. The experiment results show that, in this hybrid method, percentage of correct prediction is better than simple k-mean, and provide better recommendations to the users.

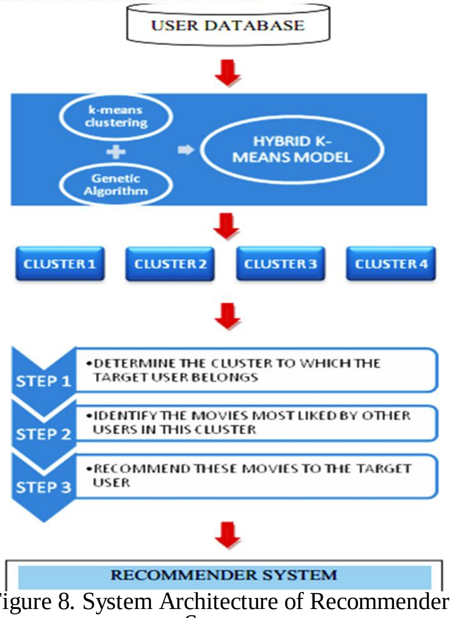

Figure 8. System Architecture of Recommender System

A method has been presented by [20] using Memetic clustering algorithm in collaborative filtering recommendations. Though traditional K-means clustering algorithms are simple and take less time in clustering, falling in local optima is more probable. Therefore, in order to overcome this problem, Memetic clustering algorithms are used in this approach. Memetic algorithms are different from Genetic algorithms. In this way, cultural evolution is constructed in Memetic algorithms by meme instead of gene, and before they get involved in evolutionary process, local search and quest has been embedded in the process to refine the solutions, so local search optimization is an elementary part of Memetic algorithms. In this approach, Collaborative filtering is performed in two phases. In the first phase, a model is extended on the basis of Memetic Clustering algorithm, and in the second phase, trained model is used to predict recommendations for active user. Experiments demonstrate that predictive accuracy of the proposed systems is clearly better than traditional collaborative filtering techniques.

Genetic and Memetic clustering algorithms are compared by [21] for collaborative recommendation. It has been concluded that accuracy of Memetic algorithm is better than the genetic algorithm in terms of prediction.

Ant based clustering is used by [22] for collaborative filtering in which recommendation process is based on the behavior of ants. In the natural world, ants release a certain amount of pheromone while walking for finding food and other ants follow a path which is rich of pheromone, and eventually find food. Since collaborative filtering is based on behavior of people with similar preferences, there is a similarity between ant behaviors and collaborative filtering. It means that recommendation is generated on the basis of similar users (ants) that are in same path (opinion). This approach contains two steps, namely, Clustering and recommendation. In clustering step, the rating matrix is clustered into K clusters through using ant based clustering, and Clusters data and centroid are stored for recommendation. Finally, Pheromone Initializing is performed for each cluster .In the second step, recommendation is generated. In this phase, at first, suitable clusters are selected on the basis of density of cluster and similarity with active user profile. Then, rating quality of items in each selected cluster is computed on the basis of average rating of the item in selected cluster and Variance of the ratings given by individual users for the item in the chosen cluster. Then, ratings of items are calculated by rating quality and average rating of the item in the chosen cluster. Later, the pheromone updating strategy is performed. It helps to increase the best solution pheromone, and in this way, the best clusters are chosen, and quality of recommendations improves. Finally TOP-N recommendation is generated for active user, and to sort the Pheromone Information for Future Experiment results of recommendations show that this approach works better for large dataset in comparison to traditional collaborative filtering.

A method has been presented by [23] combining ant based clustering and FCM for clustering users, and locating users in suitable classes and providing best clusters of users with similar concerns. The initialization is the most important part of FCM. In this approach, calculation of initial clusters' centers is based on Ant colony algorithm. In this way, the ants move initially toward each user to form heaps. The centroids of these heaps are taken as the initial cluster centers, and the FCM algorithm is used to refine these clusters. In the second stage, the users obtained from the FCM algorithm are hardened according to the maximum membership criteria to build new heaps. After clustering, the adaption distinct between active user and most similar use clusters is calculated, and then appropriate recommendation are provided for user. Ant based Algorithm helps to provide optimal solutions, and FCM considers the uncertainty in user's interests. The results show that better recommendations will be suggested through using the proposed method for clustering users.

4. CONCLUTION

Recommender systems are considered as a filtering and retrieval technique developed to alleviate the problem of information and products overload. Collaborative filtering is the most popular and successful method that recommends the item to the target user. These users have the same preferences and are interested in it in the past. Scalability is the major challenge of collaborative filtering. With regard to increasing customers and products gradually, the time consumed for finding nearest neighbor of target user or item increases, and consequently more response time is required. There are some ways to overcome this drawback such as dimension reduction algorithms, item based collaborative filtering algorithm, clustering algorithms and etc. Clustering algorithms are the most effective techniques to overcome the scalability challenge. Clustering techniques partition the user or items on the basis of rating or other features, and then finding neighbors are performed within clusters instead of within whole dataset. Too many researches have been carried out in collaborative filtering, and so many approaches have been proposed for improving scalability and recommendation accuracy by applying clustering. Various kinds of clustering algorithms have been used by researchers to resolve scalability problem in collaborative filtering. In this paper, some recent approaches have been collected and surveyed. Fuzzy clustering is one of the common clustering algorithms in which different ways have been used and applied in collaborative filtering. In this paper, some of them have been mentioned. Fuzzy clustering allows users to be member of several clusters according to their variety of interests. Due to this issue, in many cases, better results are obtained in recommendations qualities by using fuzzy clustering. Also with regard to changes occured in objective function in fuzzy clustering such as using entropy function instead of Euclidean distance or using exponential function, better results are obtained. Fuzzy clustering is improved with FP-tree if that there are many kinds of data. In this case, better results are obtained in comparison to using k-mediod clustering. Using time factor in finding nearest neighbor before fizzy clustering is one of the factors considered in resolving scalability and sparcity problem. Clustering item and user fuzzy simultaneously is one promising way to offer better recommendations. Evolutionary based clustering algorithms along with resolving local optima of classical clustering algorithms can be used in collaborative filtering and to improve scalability. Different types of these algorithms have been mentioned in this paper such as genetic, memtic and ant colony clustering. Comparing genetic, memetic and k-mean clustering algorithms results in the same data set, and it is concluded that memetic algorithm has a lower prediction error (MAE) in comparison to genetic and k-mean clustering. Ant based clustering as a common way used in collaborative filtering has better results in comparison to k-mean clustering in large datasets. Through using hybrid algorithms such as a combination of genetic and k-mean clustering or a combination of ant colony and fuzzy clustering, better results are obtained in comparison to above mentioned algorithms.

5. REFERENCES

[1] Felfering, A., Friedrich, G. and Schmidt-Thieme, L. 2007. Recommender systems. IEEE Intelligent systems,pages 18-21.

[2] Taghi, M., Khoshgoftaar, Xiaoyuan Su 2009. A Survey of Collaborative Filtering Techniques. Hindawi Publishing Corporation Advances in Artificial Intelligence. Volume 2009, Article ID pages 421-425.

[3] Christian Desrosiers, George Karypis, 2011. A comprehensive survey of neighborhood-based recommendation methods.

[4] Daniar Asanov, 2011. Algorithms and Methods in Recommender Systems. Berlin Institute of Technology, Berlin, Germany.

[5] J. Ben Schafer, Joseph Konstan, John Riedl,1999. Recommender systems in e-

commerce. Proceedings of the 1st ACM conference on Electronic commerce

[6] Matteo Matteucci, Eugenio Lentini, Paolo Cremonesi, Roberto Turrin, 2010. An evaluation Methodology for Collaborative Recommender Systems.

[7] Chuleerat Jaruskulchai, Kiatichai Treerattanapitak, 2009. Item-based Fuzzy C-Means Clustering for Collaborative Filtering. NCCIT , volume 5,issue No10, pages 30-34.

[8] Chuleerat Jaruskulchai, Kiatichai Treerattanapitak,2009. Entropy based fuzzy C-mean for item-based collaborative filtering", In Proc. the 9th Int. Symposium on Communication and Information Techonology, pages 881-886

[9] Chuleerat Jaruskulchai, Kiatichai Treerattanapitak,2010. Membership enhancement with exponential fuzzy clustering for collaborative filtering. In Proc. the 17th Int. Conf. Neural Information Processing.volume part1, pages 559-566.

[10] Chuleerat Jaruskulchai,Kiatichai Treerattanapitak, 2012. Exponential Fuzzy C-Means for Collaborative Filtering. journal of computer science and technology, volume 27, issue No3, pages 567-576.

[11] Yanhui Ma, Yonghong Xie 2010. A Novel Collaborative Filtering Framework Based on Fuzzy C-Means Clustering Using FP-Tree. International Conference on Computer Application and System Modeling. IEEE,volume12, pages 4-9.

[12] Bin Xu, Chun Chen, Deng Cai, Jiajun Bu 2012. An Exploration of Improving Collaborative Recommender Systems via User-Item Subgroups. International World Wide Web Conference Committee, france, pages 21-30.

[13] Nai-Ying Zhang, Jing Wang, Jian Yin2010. Collaborative Filtering Recommendation Based on Fuzzy Clustering of User Preferences. Seventh International Conference on Fuzzy Systems and Knowledge Discovery, volume 4, pages 1946-1950.

[14] DongYue Ling Zheng, Xinyu Zhao, Shuo Cui 2010. An Improved Personalized Recommendation Algorithm Based On Fuzzy Clustering. 3rd International Conference on Advanced Computer Theory and Engineering, volume 5, pages 408-412.

[15] Chih-Fong Tsaia, Chihli Hung, 2012. Cluster ensembles in collaborative filtering recommendation, journal of Elsevier, volume12, issue No 4, pages1417-1425.

[16] Dongting Sun, Zhigang Lou, Fuahi Zhang 2011. A novel approach for collaborative filtering to alleviate the new item cold-start problem. the 11th international symposium on communication and information technology, pages 402-406.

[17] Dongting Sun, Zhigang Lou, Gong Li 2011. A Content- Enhanced approach for cold-start problem in collaborative filtering. IEEE, pages 4501-4504.

[18] Hideyuki Mase, Hayato Ohwada 2012. Collaborative Filtering Incorporating Hybrid-Clustering Technology. International Conference on Systems and Informatics, pages 2342-2346.

[19] anupriya Bhao, Prof. Dharminder Kumar, Dr. Saroj 2011. An Evolutionary K-means Clustering Approach to Recommender Systems", International journal of Computer Science and its Applications, pages 368-372.

[20] HemaBanati, ShikhaMehta 2010. MEMETIC Collaborative filtering based Recommender System. Second International Conference on Information Technology for Real World Problems, pages 102-107.

[21] Hema Banati, Shikha Mehta 2010. a multi-perspective evaluation of MA and GA for collaborative filtering recommender system. International journal of computer science & information Technology, volume 2, issue No5, pages 103-122.

[22] H. Kaur, P. Bedi, R. Sharma 2009. Recommender System Based on Collaborative Behavior of Ants. Journal of Artificial Intelligence, pages 40-55.

[23] Ayoub Bagheri, Mohammad. H. Saraee, Mohammad Davarpanah, Shiva Nadi1 2011. FARS: Fuzzy Ant based Recommender System for Web Users. International Journal of Computer Science Issues. January,volume 8, issue No 1, pages 2-3-2009.

[24] Chenxi Lin, Gui-Rong Xue, Qiang Yang, WenSi Xi, Hua-Jun Zeng, Yong Yu,Zheng Chen 2005. Scalable Collaborative Filtering Using Cluster-based Smoothing. ACM SIGIR'05 Conf, pages114-121.

[25] Thomas George, Srujana Merugu 2005. A scalable collaborative filtering framework based on co-clustering. In Proceedings of the IEEE ICDM Conference.

[26] P. Melville, R. J. Mooney, and R. Nagarajan 2002. Content-boosted collaborative filtering for improved recommendations. in Proceedings of the 18th National Conference on Artificial Intelligence (AAAI '02), pages 187-192.

[27] M. Balabanovi, Y. Shoham, 1997. Content-based, collaborative recommendation. Communications of the ACM, volume 4,issue No 3, pages 66-72.

[28] L. Roy, R. J. Mooney 1999. Content-based book recommendation using learning for text categorization. in Proceedings of the Workshop on Recommender Systems: Algorithms and Evaluation (SIGIR '99). Berkeley. Calif, USA