

Personalized Recommender by Exploiting Domain based Expert for Enhancing Collaborative Filtering Algorithm: PReC

Mrs. M. Sridevi¹

Anurag Group of Institutions
Hyderabad, Telangana, India

Dr. R. Rajeswara Rao²

JNTU Kakinada
Vizianagaram, Andhra Pradesh, India

Abstract—The large amount of information available on the internet initiated various Recommender algorithms to act as an intermediate between number of choices and internet users. Collaborative filtering is one of the most traditional and intensively used recommendation approaches for many commercial services. Despite providing satisfying outcomes, it does have some issues that include source diversity, reliability, sparsity of data, scalability and cold start. Thus, there is a need for further improvement in the current generation of recommender system to achieve a more effective human decision support in a wide variety of applications and scenarios. Personalized Expert based collaborative filtering (PReC) approach is proposed to identify domain specific experts and the use of experts preference enhanced the performance of collaborative filtering recommender systems. A unified framework is proposed that integrates similar users rating data, experts rating and demographic data to reduce the number of pairwise computations from the search space to ensure scalability and enabled fine grained recommendations. The proposed method is evaluated using accuracy metrics MAE, RMSE on the data set collected from MovieLens datasets.

Keywords—Recommender system; collaborative filtering; domain based experts; demographic data

I. INTRODUCTION

With an overwhelming growth of information available over the internet in recent years, Recommender systems [1], [2] have proven to be a powerful tool whose aim is to guide users with personalized recommendations and alleviate the potential problem of information overload. However, the performance of these systems in dynamic environments remain unsatisfactory, thereby making Recommender systems inefficient, inaccurate and less robust to the changes of user preferences thereby motivating for further research in the field.

Despite the success of several Recommendation approaches such as Collaborative filtering [3], [4], Content based [5], [6] and Hybrid filtering [7], there have been several limitations increasing the need to provide effective and accurate recommendations. Collaborative filtering is one of the most traditional and intensively used recommendation approach for many commercial services like Movies recommendation, Music Recommendation, News Recommendation, Book Recommendation, etc., as it is

content independent and easy to implement. In this approach, recommendations are generated based on user ratings and the similarity measures between users (User-based CF) and/or items (Item-based CF). Despite providing satisfying outcomes, existing collaborative filtering methods does have some issues that include source diversity, reliability, sparsity of data, scalability and cold start. Thus, there is a need for further improvement in the current generation of recommender system to achieve a more effective human decision support, in a wide variety of applications and scenarios.

In order to address the aforementioned problems, existing solutions in the literature have been explored and an effective recommendation approach is proposed that can outperform traditional algorithms. A Personalized Expert based collaborative filtering (PReC) is proposed for domain specific expert identification by analyzing each user and item subgroups to predict relevant unseen preferences and improve the performance of the system in terms of source diversity and recommender reliability. A novel use of expert's preference elicitation enabled fine grained recommendations thereby, solving sparsity and cold start issues. Intelligent Demographic filtering mechanism is introduced to reduce the number of pairwise computations from the search space to ensure scalability. Finally, a versatile unified framework is proposed in which all the aforementioned approaches have been integrated to handle the challenges of current generation Recommender systems.

II. RELATED WORK

Recommender system has come into existence to overcome the choice option by providing recommendations that are tailored to individual user preferences. Research shows that Recommender systems [1] [2] have increased the sales and customer satisfaction. It guides users by providing personalized information in a large space of options. Collection of input data is one of the important components of Recommender system. It considers user preferences in the form of explicit rating i.e., providing 5-star ratings or implicit data by clicking, browsing, etc. Since the inception of Recommender systems, today we have many Recommendation strategies/algorithms based on the user ratings, content of the item, location, content, etc.

A. Collaborative Filtering

Collaborative filtering [3] [4] is one of the widely used technique to predict users preference or generate list of user preferences. However, there exists several issues which need to be addressed that include the complex computation for large dataset of users, understand the diversity of different information sources and understanding user preferences when there is sparsity of data. Prior research in Collaborative Filtering Recommender systems focus only on the accuracy of the recommendations at the cost of scalable and reliable recommendations. Among the various Recommender systems used, early systems that used collaborative filtering approaches are Grundy System a book recommender, Ringo [8] which recommended personalized music, Tapestry [9], a document recommender, an article recommender GroupLens [10], Amazon.com [11] recommender which recommends the relevant items, etc., In [12] improved collaborative filtering approach is used by extending user-item rating matrix that could alleviate the data sparsity issue but could not solve cold start problem and user interest shifted with time. In [13] Regularized Matrix Factorization is used to develop an Incremental Collaborative Filtering Recommender system.

B. Content-Based Recommender System

Content based recommender systems [5], [6] can be used in a wide variety of domains ranging from recommending news articles, hotels, movies and items for sale. Also referred as cognitive filtering, the key idea of content based recommender is to recommend items to the user according to relevant attributes of the item and user preferences by attributes. Content based recommender system [6] compares the items to others in the collection. Items with a high degree of similarity are presented as recommendations. This is called as “item-to-item correlation”. The information source used by the content-based filtering systems is text documents. These items are typically described with keywords and weights. By analyzing the keywords and document the system recommend a suitable item using probabilistic methods, for e.g., Naïve Bayes, decision trees [14], clustering, nearest neighbor, etc., There are many ways to build a user profile which involves (i) inferring profile from user actions (implicit). It includes read, buy, click, etc. and (ii) inferring profile from explicit user ratings, which includes feedback technique by selecting a value of range, filling out forms, etc., Both implicit actions and explicit ratings are processed against the database of content attributes to build the content profile. Content based recommender systems have been applied in various applications. The Fab system [5], which recommends Web pages to users, [15] represents documents, DailyLearner system [16] filter out news items too similar to those already seen by the user, etc. Content based recommender is effective in recommending accurate items when they are better described. The profiles of other users do not influence the recommendations of the target user as the recommendations are based on individual preferences. Content-based recommender systems always need to consider the following issues:

1) *Partial content analysis*: It is difficult to recommend about the user’s profile if only partial content available.

2) *High specialization*: It restricts the users to rate the items by identifying the similar items defined in their corresponding profiles and so new items and other options are not revealed to rate.

C. Hybrid Recommender System

Hybrid Recommender system [17] combines both collaborative filtering and content based filtering approach in different ways: i) Implementation of algorithms separately and combining the results [18]; ii) Utilizing some capabilities of content based into collaborative based approach [19]; iii) Utilizing some capabilities of collaborative based into content based approach [20]; and iv) Constructing a unifying model using different approaches [21].

Hybrid Recommender system integrates various RSs which helps us to overcome the disadvantages and improve the performance. Therefore the quality of the recommendations provided to the end user or customer. Empirically, several studies compared the performance and quality of hybrid approach [22] and proved that the hybrid approaches provide recommendations more accurately than original recommender approaches. The very common issues cold start and sparsity are also eliminated.

D. Expert based Recommender Systems

A series of algorithms have been described in the literature to identify the experts and exploit their opinion to demonstrate their effectiveness. Earlier research expert is judged by comparing with a set of predefined keywords [22], that is time-consuming and also the profile becomes outdated that no longer reflects in future. This resulted an increased demand for automating and more focused towards the development of personalized expert based Recommender systems. To capture the expert, a number of approaches exist in the literature, like Probabilistic models [23] that can estimate the association between domain and expert users, discriminative models [24] that directly estimate the conditional probability to find the relevance, voting models [25] use voting mechanism to vote the users, graph based models [26] determine the associations by inference on a graph consisting items, users, etc., and there are some other models that make use of latent variables that correspond to a theme. In [27], expert ranking is considered as a voting problem using data fusion techniques. Author in [28] used context dependent expertise information to reduce the number of users. In [29], Ranking SVM is applied to rank the experts by using pairwise approach to rank and predict the candidates. In [30] an evaluation of Learning to Rank algorithms is proposed for expert search on the DBLP database. In [31] a supervised learning approach is proposed to aggregate ranking and apply the same to search the experts and their blogs. An expert user is identified through the blog entries in [32] to generate recommendations. Experts identified from the user community dynamically proved to be time consuming in [33].

E. Demographic based Recommender System

Demographic Recommender system [34] considers demographic data in collaborative filtering when providing recommendations. In [35], recommendations are generated based on the product demographic data learned from online

reviews and blog entries. The demographic features of the tourist are taken in [36] without any rating data to generate the predictions but it can only achieve limited accuracy. A framework is developed by evaluating demographic attributes to solve cold start problem in [37].

III. PROPOSED METHODOLOGY

In this paper, we proposed personalized expert based collaborative filtering (PReC) to identify domain specific experts and use of demographic data with expert's preference in order to improve the performance of traditional collaborative filtering recommender systems. Initially, EUCF is proposed that integrates collaborative filtering features and clusters similar users and similar items thereby promoting experts based on the user profile and exploit their opinions thus addressing the reliability issue and enables fine grained recommendations. Secondly, an intelligent demographic filtering mechanism is introduced that integrates demographic features and user based collaborative filtering to reduce the number of pairwise expensive similarity computations from the search space to ensure scalability and cold start issue. And finally, a unified framework is proposed that integrates similar users rating data, experts rating and demographic data to deal source diversity, sparsity and enable fine grained accurate recommendations.

A. Expert user based Collaborative Filtering (EUCF)

Based on the study, we developed a framework as given in Fig. 1 to identify the experts from users' responses having similar preferences to improve recommendation quality. While traditional systems compare users profile with other users to recommend unseen movies, our proposed system compare expert users profile with target user and use the experts' opinion that eventually results in better recommendations to the target user.

1) *Neighborhood formation:* Once the rating profile of the user is constructed, similar users are identified by analyzing user profile (User-Item Rating Matrix) and by applying the similarity measure, Pearson correlation coefficient, equation (1).

$$s(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

where x_i, y_i is the rating of user x and y on 'i', \bar{x} mean of all x 's ratings, \bar{y} , mean of all y 's ratings and n is the number of items

Likewise, the similarity between items is computed by using the user item ratings and discovering similar items. For computation of similarity between items, Pearson correlation coefficient is used as given in equation (2).

$$s(t, r) = \frac{\sum_{i=1}^n (R_{it} - \bar{A}_t)(R_{ir} - \bar{A}_r)}{\sqrt{\sum_{i=1}^n (R_{it} - \bar{A}_t)^2} \sqrt{\sum_{i=1}^n (R_{ir} - \bar{A}_r)^2}} \quad (2)$$

where R_{it}, R_{ir} is the rating of user i on item 't' and item 'r' respectively, \bar{A}_t the average rating of item 't', \bar{A}_r , average rating of item 'r' and n is the number of users.

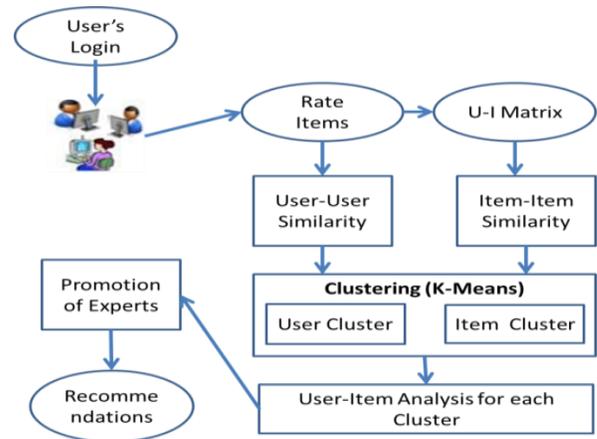


Fig. 1. Framework of EUCF.

2) *Expert identification and generating recommendations:* Designing a framework for expert identification involves collecting user rating data from user – item profile that contributes in identifying users with highest access factor and most similar to the average ratings opinion in exchange of high quality personalization. A pseudo rating profile is generated by taking the average ratings of the users in that cluster.

Once the identification of similar users and similar items is completed, the selection of the right expert during identification phase is critical. Expert user is a user who is most similar to the average ratings and whose access factor is greater than the other users. An expert user may vary from one domain to another domain and there may exist more than one expert user for the same domain. The task of expert finding usually involves taking the user rating data as input, denoting the heavy access factor and finding similarity with average ratings and returning the list of users ordered by their expertise level. User's expertise is characterized by the heavy access factor i.e., more number of ratings for a given domain and is given in equation (3).

$$\forall_{u \in U} \forall_{i \in I} i_{HA}^{(u)} = \frac{\log(|I_u|+1)}{\max_v \log(|I_v|+1)} \quad (3)$$

Where I_u set of items 'u' accessed, $HA = 1$ if 'u' has heavy access. The features of user profile related to the number of items rated associated with each domain and also identifying the user profile who is most similar to the average rating assuming that the users with such profiles are likely to be considered as experts. The average rating similarity is calculated as shown in equation (4).

$$\forall_{u \in U} \forall_{i \in I} d(u) = \sqrt{\sum_{i \in I} \sum_{i=1}^n \left((R_{u,i}) - \bar{R}_{U,i} \right)^2} \quad (4)$$

Where U, I are the user set and item sets respectively, n is the number of items, $\bar{R}_{U,i}$ average rating of users who rated item 'i'. Thus, the expertise can be measured by the following equation (5).

$$E(u) = i_{HA}^{(u)} * \left(\frac{1}{1+d(u)} \right) \quad (5)$$

In selecting the experts, threshold value of fixed size can be defined. Small threshold value often results in high precision but low recall in experts finding, whereas large threshold value results in low precision but high recall. We assume that there may exist more than one candidate expert for every domain. Formally, given the user ratings, candidate experts are identified and the similarity of expert user and the target user is computed as shown in equation (6).

$$s(e, a) = \frac{\sum_{i=1}^n (R_{e,i} - R_{\bar{e}})(R_{a,i} - \bar{R}_a)}{\sqrt{\sum_{i=1}^n (R_{e,i} - R_{\bar{e}})^2} \sqrt{\sum_{i=1}^n (R_{a,i} - \bar{R}_a)^2}} \quad (6)$$

and then finally predictions are generated based on similar experts recommendations. The expert scores are averaged to generate predictions by the following equation:

$$E(i) = \hat{R}_a + \frac{\sum_{e=1}^n (R_{e,i} - \hat{R}_e) s(e, a)}{\sum_{e=1}^n s(e, a)} \quad (7)$$

Where \hat{R}_a average rating of target user, $R_{e,i}$ rating of expert user on target item, \hat{R}_e average rating of target user to the items, $s(u, e)$ similarity of target user 'u' and expert user 'e'. An expert candidate is assessed by considering how many items (movies) have been rated, how many are related to the particular domain, and how similar the user with the average opinions.

B. Demographic User based Collaborative Filtering (UCFD)

We studied the difficulty in computing the similarity among different users from a large dataset. A novel approach is proposed for efficient computations of similar users based on demographic data of the user. We use demographic features of users to partition the set of users having similar demographic features i.e., age and gender and from the subset again, by analyzing the ratings of the user, similar users are identified by their ratings preferences. We assume that users with similar demographic features have similar user preferences that will effectively improve the time and performance in traditional Collaborative filtering system.

Thus a unified framework is proposed that utilizes both rating data and demographic data to compute the similarity between users thus bypasses the scalability issue. Formally, the profile of a user is represented as a vector where the elements of the vector correspond to the demographic features of the user as shown in Table I. The input contains a demographic feature vector space of each single user and the output is the relevant similarity score of the users. By selecting a set of well-designed features i.e., age and gender, our proposed approach outperforms the traditional models in terms of scalability on the dataset extracted from MovieLens dataset. After data selection, pairwise computations are performed to partition the users into clusters. The demographic filtering of users mainly reduces the pairwise similarity computations that are performed during user similarity computation. Our proposed approach has the advantage of overcoming the sparsity when there is sparsity of data. The framework combines demographic based and user based collaborative filtering recommendations in order to derive benefits.

TABLE I. DEMOGRAPHIC DATA

Attributes	Demographic vector description
age<=18	slot takes value 0
18<age<=35	slot takes value 1
36<age<=49	slot takes value 2
age>50	slot takes value 3
Male	slot describing gender takes 1
Female	slot takes value 0

Similarity of users based on demographic data can be computed using the following equation:

$$\text{dem}_{u,a} = \cos(u, a) = \frac{\vec{u} \cdot \vec{a}}{\|\vec{u}\| \|\vec{a}\|} = \frac{\sum_{i=1}^n u_i a_i}{\sqrt{\sum_{i=1}^n u_i^2} \sqrt{\sum_{i=1}^n a_i^2}} \quad (8)$$

and the similarity based on user rating data of similar demographic users is computed using Euclidean distance measure as shown in equation (9):

$$d(u, a) = \sqrt{\sum_{i=1}^n (R_{u,i} - R_{a,i})^2} \quad (9)$$

where, $R_{u,i}, R_{a,i}$ rating of user u and a on item 'i' and 'n' is the number of items and the similarity between the users is given by the equation (10).

$$s(u, a) = \frac{1}{1+d(u, a)} \quad (10)$$

Finally, as a result our proposed enhanced correlation is based on demographic correlation and similarity rating based correlation given by the following equation (11):

$$\text{dem_rat_sim}(u, a) = \alpha \text{dem}_{u,a} + \beta s_{u,a} + \gamma (\text{dem}_{u,a} * s_{u,a}) \quad (11)$$

UCFD generates prediction for the target user based on the enhanced correlation given by

$$\text{Pred_dem_rat_sim}(a, i) = \hat{R}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * \text{dem_rat_sim}(u, a)}{\sum_{u=1}^n \text{dem_rat_sim}(u, a)} \quad (12)$$

where $\text{Pred_dem_rat_sim}(a, i)$ predicted rating for target user a for item i, $\text{dem_rat_sim}(u, a)$ similarity between user u and a, \hat{R}_a is average ratings of target user, $r_{u,i}$ rating of user u on item i, \bar{r}_u average ratings of user u and n is the number of similar neighbors.

C. Expert user based CF with Demographic Filtering (EUCFD)

A novel recommendation algorithm is proposed that fuses the opinions from similar users based on rating data, expert users (EUCF), and similar users based on demographic features (UCFD). A typical recommender system considers only one's judgment i.e., similar user, whereas our proposed system considers two more sources i.e., Expert users and demographic similar users in addition to similar users predictions to generate efficient output as shown in Fig. 2.

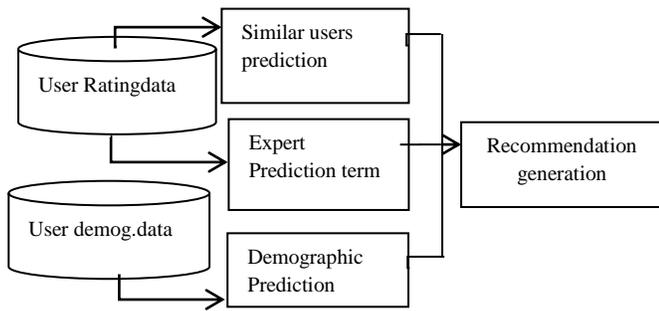


Fig. 2. Architecture of Proposed EUCFD.

The EUCF method extracts expert users from each item category and exploits their ratings rather than similar neighbors. In addition, the similarity between the target user and the users is considered based on demographic features in the prediction of ratings when there is sparsity of rating data and cold start issue and finally we fuse all the variants i.e., EUCF and UCFD to form EUCFD that utilizes domain based experts opinion by considering both demographic and ratings data.

Initially, we identify expert users from an independent user data set based on the number of ratings given to each item category. The demographic features of the active user are considered if there is sparsity of rating data. Recommendations are generated by considering the prediction terms of expert user and the accumulative values of similar users based on demographic data and similar users rating data.

1) *Enhanced prediction term for target user:* Finally, given a user's choice of preference, personalized recommendations are generated based on their preferences from the movies space. The algorithmic approaches are ensemble to generate recommendations by considering three varying factors: similar users prediction term, expert based prediction term, demographic based prediction term as given in the following equations:

$$SP(a,i) = \hat{R}_a + \frac{\sum_{u=1}^n (R_{u,i} - \bar{R}_u) * s(u,a)}{\sum_{u=1}^n s(u,a)} \quad (13)$$

$$EP(a,i) = \hat{R}_a + \frac{\sum_{e=1}^n (R_{e,i} - \bar{R}_e) * s(e,a)}{\sum_{e=1}^n s(e,a)} \quad (14)$$

$$DP(a,i) = \hat{R}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * dem(u,a)}{\sum_{u=1}^n dem(u,a)} \quad (15)$$

From equation (13)(14)(15), we compute enhanced prediction term for target user as:

$$P(a,i) = \bar{R}_a + \alpha * SP(a,i) + \beta * EP(a,i) + \gamma * DP(a,i) \quad (16)$$

Where $\alpha + \beta + \gamma = 1$, \bar{R}_a is the average of target user ratings, $SP(a,i)$ is the user similarity prediction term, $EP(a,i)$ is the expert users prediction term and $DP(a,i)$ is demographic prediction term. Parameter α, β, γ is used to tune between the three values. We aggregate the recommendations generated from each module by considering the weights. When there is an absence in the explicit user input ratings, thus resulting in the sparsity of data thereby resulting inaccurate recommendations, then the canonical proposed approach is to consider the demographic features of the users retrieved from

the user profile and use these features to find the similarity between the users and to set recommendations for the target user.

D. Experimental Evaluation

1) *Datasets:* In this study, we considered MovieLens datasets. MovieLens dataset consists of approximately 100,000 ratings, 943 users and 1682 movies, where each user has rated atleast 20 movies and the ratings are on a 1-5 scale with '1' representing least and '5' representing highest. MovieLens is a web based Recommender system. The data used by Movie Lens is collected by GroupLens Research Project at Minnesota University. (<http://grouplens.org/datasets/movielens/100k/>). The statistics of the datasets are detailed in Table II.

2) *Evaluation metrics:* In order to evaluate our system, the data is divided into two sets with 80% - 20% split ratio of training set and testing set. Let 'x' be a variable that give the percentage of training and test data. If x=0.8, then it indicates that 80% of data is used as training set and 20% of data is used as test set. The performance of the proposed system can be evaluated by employing Recommendation Accuracy Metrics. In this paper, we use Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) most frequently used Metrics to measure the accuracy.

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - r_i| \quad (17)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (p_i - r_i)^2}{n}} \quad (18)$$

Where p_i prediction ratings for item 'i', r_i actual ratings for item 'i' and n is number of rated items. The lower the MAE value, the more accurate are the recommendations. The other metrics used in this study are precision and recall.

$$\text{Precision} = \frac{|\text{relevant_items_recommended_items}|}{|\text{recommended_items}|} \quad (19)$$

$$\text{Recall} = \frac{|\text{relevant_items_recommended_items}|}{|\text{relevant_items}|} \quad (20)$$

There is a certain ambiguity exists while using these measures. For an instance, increasing the total number of recommended items N, which directly effects on increase in recall but decreases precision. To overcome this ambiguity, we use another metric F1-Measure. It gives equal weightage to both precision and recall. All these measures are defined in equations (19), (20) and (21) individually.

$$F1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (21)$$

TABLE II. DATASET STATISTICS

Statistics	MovieLens
No.of Users	943
No.of Movies	1682
No.of Ratings	100,000
Min-Max values	1 - 5

3) *Experimental results:* Finally, we evaluate our results and experiments on MovieLens datasets. The experiments proved that our proposed algorithms generate efficient and accurate predictions when compared to existing traditional User based Collaborative Filtering. The Mean Absolute Error of our proposed algorithms is lower than that of UCF. The outcomes from all the proposed algorithms are analyzed and

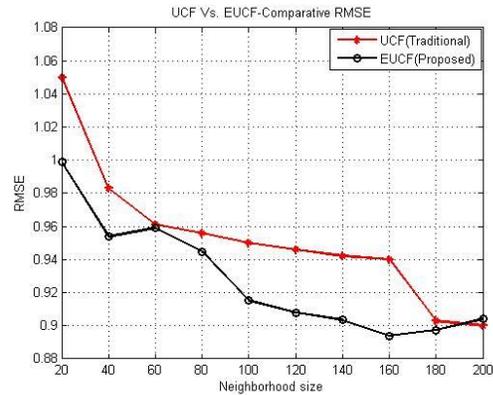
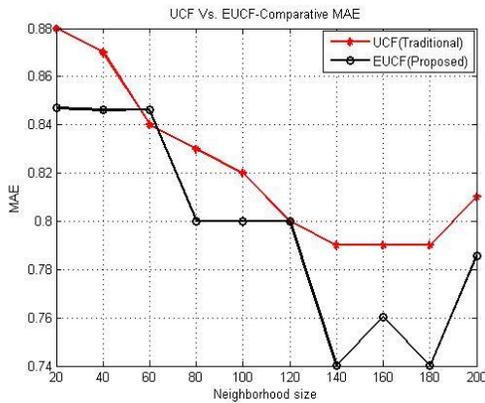
compared based on MAE. Table III gives the comparison of MAE values for all the proposed algorithms for different sizes of neighborhood. In order to make the improvement even prominent, the results are also compared for root mean square error formats given in Table IV. The outcomes are also analyzed graphically in Fig. 3, Fig. 4 and Fig. 5.

TABLE III. MAE VALUES OF PROPOSED ALGORITHMS BASED ON NEIGHBORHOOD SIZE

Algorithm	Neighborhood size									
	20	40	60	80	100	120	140	160	180	200
UCF	0.88	0.87	0.84	0.83	0.82	0.8	0.79	0.79	0.79	0.81
EUCF	0.8469	0.8462	0.8463	0.8001	0.8001	0.8001	0.7401	0.7601	0.7401	0.7857
UCFD	0.7381	0.7381	0.7381	0.6578	0.6595	0.6588	0.5693	0.5693	0.5693	0.6386
EUCFD	0.716	0.7254	0.7071	0.6025	0.5957	0.5689	0.5186	0.5061	0.5494	0.5498

TABLE IV. RMSE VALUES OF PROPOSED ALGORITHMS BASED ON NEIGHBORHOOD SIZE

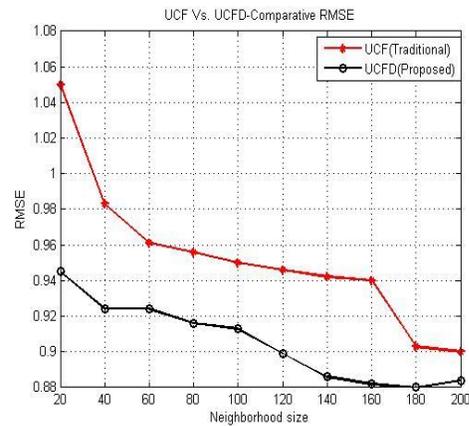
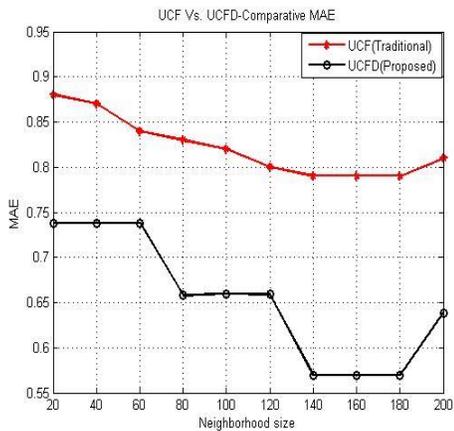
Algorithm	Neighborhood size									
	20	40	60	80	100	120	140	160	180	200
UCF	0.7744	0.7569	0.7056	0.6889	0.6724	0.64	0.6241	0.6241	0.6241	0.6561
EUCF	0.7172	0.716	0.716	0.6401	0.6401	0.6401	0.5477	0.5777	0.5477	0.6173
UCFD	0.5448	0.5448	0.5448	0.4327	0.435	0.4341	0.3241	0.3241	0.3241	0.407
EUCFD	0.512	0.5263	0.5	0.363	0.3549	0.323	0.269	0.256	0.301	0.302



(a) UCF vs. EUCF- MAE

(b) UCF vs. EUCF- RMSE

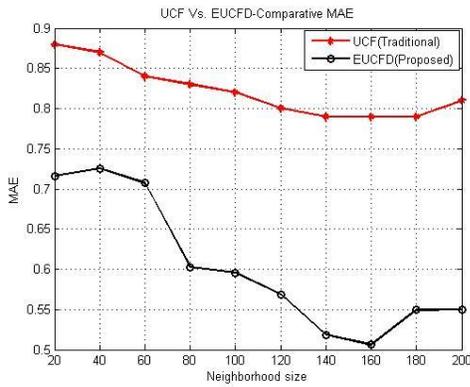
Fig. 3. Performance vs. Neighborhood Size, MovieLens Dataset.



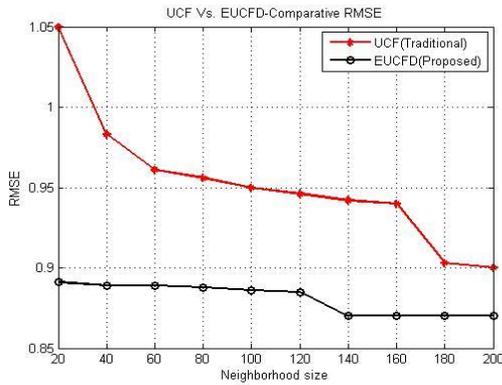
(a) UCF vs. UCFD- MAE

(b) UCF vs. UCFD- RMSE

Fig. 4. Performance vs. Neighborhood Size, MovieLens Dataset.



(a) UCF vs EUCFD MAE



(b) UCF vs. EUCFD RMSE

Fig. 5. Performance vs. Neighborhood Size, MovieLens Dataset.

We compared all three algorithms with the traditional algorithm and is summarized in Table V. It has been observed from Fig. 6 that EUCFD is more accurate than traditional user based collaborative filtering (UCF), item based collaborative filtering (ICF) and proposed UCFD, EUCF algorithms and the results proved that our proposed EUCFD algorithm outperformed remaining algorithms and our proposed method shows a significant increase in prediction performance when compared to a traditional single source model.

The results in Fig. 7 demonstrates the higher relevancy of the final EUCFD algorithm recommendations and the significant improvements made over UCF, EUCF and UCFD algorithms for MovieLens datasets. The F-Measure results of EUCF, UCFD and EUCFD are also analyzed graphically as well.

TABLE V. SUMMARY OF MAE AND RMSE (MOVIELENS)

S. No.	Algorithm	MAE	RMSE
1	User based CF (UCF)	0.8220	0.739
2	Item based CF (ICF)	0.8154	0.711
2	Expert user based CF(EUCF)	0.7966	0.636
3	Demographic user based CF (UCFD)	0.6537	0.432
4	Expert user based Demographic CF(EUCFD)	0.6040	0.372

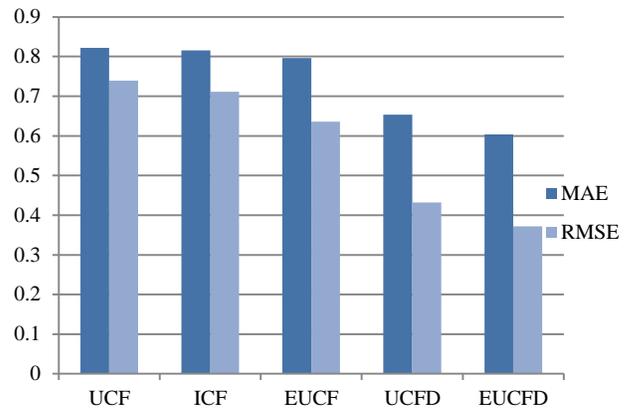
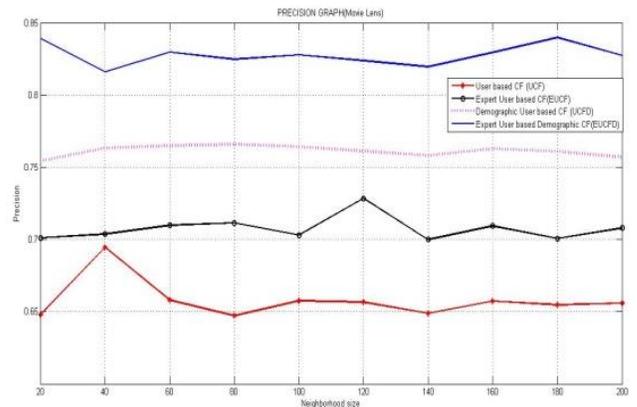
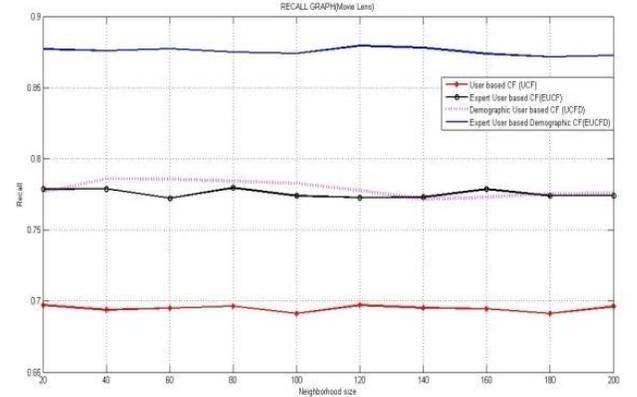


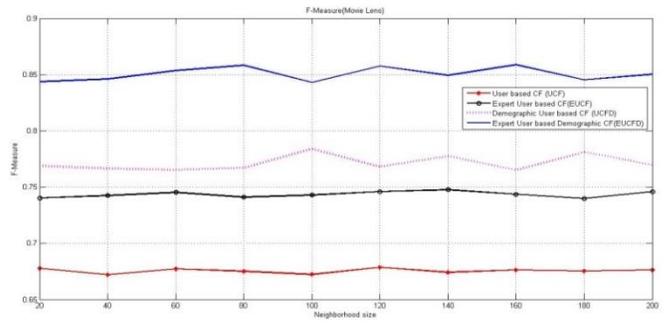
Fig. 6. MAE of UCF, EUCF, UCFD and EUCFD in MovieLens.



a) Precision



b) Recall



c) F Measure

Fig. 7. Performance vs. Neighborhood Size, MovieLens Data Set.

IV. CONCLUSION

This paper introduced the benefits to the users of information systems in retrieving their personalized preferences by considering rating data, experts' opinion and demographic data of the user to generate personalized preferences. Furthermore, we introduced novel methods in identifying domain based experts and integrated similar users' opinion with experts' opinion to improve the recommendation quality. This paper focus on the issues of Collaborative filtering that includes data sparsity, scalability and reliability. Initially, we contributed three approaches that significantly improved the recommendation quality. Results proved that our approach can increase the scalability and generate an accurate prediction that is more suitable for large data sets. Thus, our proposed approach works towards the development of an efficient recommender system.

REFERENCES

- [1] Adomavicius G., Tuzhilin, A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, *IEEE TKDE*, 17(6), 2005, pp.734-749.
- [2] JoeranBeel, BelaGipp, Stefan Langer, and CorinnaBreitinger"Research-Paper Recommender Systems: A Literature Survey", *International Journal on Digital libraries*, 2015.
- [3] Yue Shi, Martha Larson, And Alan Hanjalic Collaborative Filtering beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges, *ACM*, 2014.
- [4] J.B.Schafer,D.Frankowski,J.Herlocker,S.Sen, Collaborative filtering recommender systems. *The Adaptive Web*, 291-324, 2007.
- [5] Pasquale Lops, Marco de Gemmis and Giovanni Semeraro ' Content based Recommender systems: State of the art and trends', *Recommender Systems Handbook*, Springer Science+Business Media, LLC 2011.
- [6] Balabanović, M. and Shoham, Y. Fab: content-based, collaborative recommendation. *Communications of the ACM* 40 (3) (1997) 66-72.
- [7] Robin Burke. Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User Adapted*.
- [8] U. Shardanand, P. Maes. Social information filtering: algorithms for automating word of mouth". *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 210-217, (1995).
- [9] D. Goldberg, D. Nichols, B. Oki and D. Terry. "Using Collaborative Filtering to Weave an Information Tapestry." in *ACM*. New York. 1992.
- [10] P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl,"GroupLens: An Open Architecture for Collaborative Filtering of Netnews," *Proc. 1994 Computer Supported Collaborative Work Conf.*,pp. 175-186, 1994.
- [11] G. Linden, B. Smith, J. York. Amazon.com Recommendations, Item-to-Item Collaborative Filtering. *IEEE Internet Computing*, Volume 7, Issue 1, (2003).
- [12] Li Zhang, Tao Qin PiQiangTeng, "An Improved Collaborative Filtering Algorithm Based on User Interest" *Journal of Software*, 2014.
- [13] Luo, X., Xia, Y., and Zhu, Q, Incremental Collaborative Filtering Recommender based on Regularized Matrix Factorization, knowledge based systems,2012.
- [14] L. M. de Campos, J. M. Fern_andez-Luna, J. F. Huete, and M. A. Rueda-Morales, Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks," *International Journal of approximate Reasoning*, vol. 51,pp. 785{799, Sept. 2010.
- [15] Degemmis, M., Lops, P., Semeraro, G.: A content-collaborative recommender that exploits wordnet-based user profiles for neighborhood formation. *User Modeling and User-Adapted Interaction*17(3), 217–255 (2007).
- [16] Billsus, D. and M. Pazzani. User modeling for adaptive news access. *User Modeling and User-Adapted Interaction*, 10(2-3):147-180, 2000.
- [17] Wenfulin, Yingli,Shuangfeng, Yongbin Wang," The optimization of weights in Weighted Hybrid Recommendation algorithm, *ICIS/IEEE*, 2014.
- [18] M GhazanfarAPrugel-Bennett "An Improved Switching Hybrid Recommender System Using Naive Bayes Classifier and Collaborative Filtering- 2010.
- [19] L. M. de Campos, J. M. Fern_andez-Luna, J. F. Huete, and M. A. Rueda-Morales, Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks," *International Journal of approximate Reasoning*, vol. 51,pp. 785{799, Sept. 2010.
- [20] M. Tiemann and S. Pauws, (Towards ensemble learning for hybrid music recommendation," in *Proceedings of the 2007 ACM conference on Recommender systems, RecSys '07*, (New York, NY,USA), pp. 177{178, ACM, 2007.
- [21] Thomas HornungCai-Nicolas Ziegler Simon Franz Martin Przyjaciel-Zablock Alexander Schatzle Georg Lausen "Evaluating Hybrid Music Recommender Systems" *IEEE* 2013.
- [22] M.J. Pazzani, A Framework for Collaborative, Content-based and Demographic Filtering. *Artificial Intelligence Review*, 13(5-6), 393-408, 1999.
- [23] Amatrian, X., Lathia, N., Pujol, J.M., Kwak, H., Oliver, N.: The Wisdom of The Few: A Collaborative Filtering Approach Based On Expert Opinions From The Web. In: *32nd International Conference ACM Special Interest Group on Information Retrieval*, New York, pp. 532–539 (2009).
- [24] Kumar, A. and Bhatia, M.P.S. Community expert based recommendation for solving first rater problem. *International Journal of Computer Applications* 37 (10) (2012) 7-13.
- [25] Craig Macdonald and IadhOunis, "Learning Models for Ranking Aggregates" *LNCIS, Springer*, Vol. 6611, 2011.
- [26] Zhou Zhao, Lijun Zhang, Xiaofei He, and Wilfred Ng "Expert Finding for Question Answering via Graph Regularized Matrix Completion", *IEEE Transactions On Knowledge And Data Engineering*, Vol. 27, No. 4, April 2015.
- [27] Song, S., Lee, S., Park, S., Lee, S.G.: Determining User Expertise For Improving Recommendation Performance. In: *6th International Conference on Ubiquitous Information and Communication 2012*, New York (2012).
- [28] V.Carchiolo, A.Longhue, M.Malgeri,G.Mangioni Searching for experts in a context-aware recommendation network, *Elsevier*, Vol.51, 2015
- [29] Catarina Moreira, PavelCalado, Bruno Martins "Learning to Rank Academic Experts in the DBLP Dataset", *Expert systems: Journal of Knowledge Engineering archive*, Vol 32, Issue 4 August 2015.Pg. 477–493.
- [30] W.S.Hwang, Ho-Jong Lee,S.W.Kim, Y.Won, M.Lee, Efficient recommendation methods using category experts for a large dataset", *Information fusion*, Elsevier, 2016.
- [31] Kibeom Lee and KyoguLee,"Using Dynamically Promoted Experts for Music Recommendation", *IEEE*, 2014.
- [32] K. Lee and K. Lee, "Using experts among users for novel movie recommendations" *Journal of Computer Science & Engineering*, 2013.
- [33] Yi Fang, LuoSi, Aditya "Discriminative probabilistic models for expert search in heterogeneous information sources" *ACM* 2011.
- [34] Gupta, J. and Gadge, J. Performance analysis of recommendation system based on collaborative filtering and demographics. *IEEE International Conference on Communication, Information & Computing Technology (ICCICT)*, 2015, 1-6.
- [35] Wayne Xin Zhao, YanweiGuo, Yulan He, Han Jiang, Yuexin Wu and Xiaoming Li," we know what you want to buy: A Demographic-based System for Product Recommendation on Microblogs", *ACM*, 2014.
- [36] Yuanyuan Wang,Stephen Chi-Fai Chan,Grace Ngai"Applicability of Demographic Recommender System to Tourist Attractions: A Case Study on Trip Advisor", *IEEE* 2012.
- [37] LailaSafoury and Akram Salah" Exploiting User Demographic Attributes for Solving Cold-Start Problem in Recommender System", *Lecture Notes on Software Engineering*, Vol. 1, No. 3, August 2013.