A Proposed Method to Solve Cold Start Problem using Fuzzy user-based Clustering

Syed Badar Ud Duja[1], Baoning Niu[2]*, Bilal Ahmed[3], M. Umar Farooq Alvi[4]
Muhammad Amjad[5], Usman Ali[6], Zia Ur Rehman[7], Waqar Hussain[8]

College of Information and Computer, Taiyuan University of Technology, Taiyuan, Shanxi, China

Abstract—With the elevation of the online accessibility to almost everything, many logics, systems and algorithms have to be revised to match the pace of the trends among the socialized networks. One such system; recommendation system has become very important as far as the socialized networks are concerned . In such paced and vibrant environment of the online accessibility and availability to heavy and large amount of data uploaded to the internet such as, movies, books, research articles and much more. The method of recommendation where provides the socialized networks between the operators, at the same instance, it provides references for the users to asses other users that effects their socialized relation directly or indirectly. Collaborative filtering is the technique used for recommending the same taste of picks to that of the user, and it is accomplished by the user's mutual collaboration, this technique is mostly used by the social networking sites. Nowadays this technique is not only popular but common for recommending the data to the user; meanwhile it also motivates the researchers to find the more effective system and algorithm so that the user's satisfaction can be achieved by recommending them the data according to their search history. This paper suggests the CF (Collaborative Filtering) model that is based on the user's truthful information applied by the FCM (Fuzzy C-means) clustering. This study proposes that the fuzzy truthful information of the user is to be combined with rating of the content by other users to produce a recommender system formula with a coupled coefficient with new parameters. To achieve the results the Data set of Movie Lens is included in the study which shows significant improvement in the recommendation subjected to the condition of cold start.

Keywords—*Recommender system; collaborating filtering; cold start problem; clustering; user based clustering*

I. INTRODUCTION

Commercial e-business and e-commerce websites or social media networking websites can depicted as the instance of recommender system. These kind of online attractions use recommender systems to suggest the likewise articles and content to the viewers. It is used to filter the information available on the internet related to any particular content to suggest that content to the user. Seeking that a likewise content that has been rated or watched frequently by the similar type of users is to be recommended to other users with same taste. The historical interests of the user plays an important role in recommending the content based on the rating and preferences of other users. Recommendation systems are used in the variety of areas like in recommending videos on YouTube, shopping products on Amazon and similar apps on play stores. There are several approaches in use to recommend particular

articles to the user such as collaborative filtering, content-based recommendation or personality based recommendation and knowledge based recommendation system. All these approaches are in use depending upon the necessity of the platform for which it is being used. To recommend the content to the user, different types of engines, components and elements are fed by the recommender systems. These components normally consists of data collectors, whose job is to collect the likewise data available on the internet based on the keyword used for the search or other web analytics tools. The second component is data processor which processes the collected data for the third component that is recommender model. Which then processed for the user interface and again finally to the last component that is processing of the data that is recommended by the recommender by the model. These all components take their input from the first clue is given on the platform from the user[1][2]. Thus, such tools that recommends content to the user or group of users that supports them to explore the content they like available on the internet from their interests taken from their search history that has been rated or watched by the other users in that group to let them watch or go through the explicit set of content. Online item recommendation is achieved by the collaborative filtering technique. The past activities performed by the user such as conducting a purchase or selected by that user or other users, are given the numerical values called rating of the item, worked as the aid for the collaborative recommender model to filter out and suggest the content [3]. This technique has been used in many applications and platforms, showing very efficient results in certain environments. Collaborative Filtering recommender system at first step uses rating of the item by judging the similarities between the users rated that item. Thus, this recommender system worked on the assumption that users who liked this item previously will like the recommended content and if this particular user likes this content then the item recommended will also be liked by the users of same interest. Hence, in this system items recommended to the other users is based on the computation of similarities between users and rating of the content by other users. There are many statistics similarity measures can be implemented to determine the similarity of the rating and other statistics of the users. Pearson's correlation is one such statistics approach to determine the suggestions for the recommendation, for which the value can be somewhere between 1 and -1 to determine the extent to which two variables are linearly related to each other. Similarly, cosine similarity measures the similarity between two vectors and degree between them in the inner space of the product. These

*Corresponding Author.

both techniques are themselves interrelated to each other in the context of mean deduction, but dissimilar when it comes to the measure of coefficients [4].

The generation of the top recommendations the most popular problem i.e. cold start problem which is most popular and commonly faced challenge. Different methodologies by the researchers have been proposed to address the problem of cold start. The occurrence of this problem is faced much when there is not much data is available for particular item or for particular user. Basically, it is the problem in the information system to filter out the data available on the internet based on the data like rating of the item and interest of the user, and this data is not sufficient to infer any results for the recommendation system model. In this study it refers to the inadequate interactions between the users and the architecture of the recommended model and the quality of the recommendations are degraded significantly. In addition this study is focused on the rating of the user items to find in the nearest neighbors.

The mentioned problem then resolved by using the one of collaborative filtering technique. Collaborative filtering techniques as mentioned earlier is achieved on two bases that are Item based collaborative filtering and user based collaborative filtering, the techniques are used to recommend items from the action of other or same user respectively. These both techniques in practice are very effective but when it comes to the dynamic environment of the users or group of users, then these techniques are hard to become scalable subjected to the condition that items does not change too much. However, item based collaborative filtering approach is easy to compute offline and re-training is not required that is why KNN technique is the right thing to go for recommendation model. It also provides foundation to the development of recommendation system. However the KNN's learning method is lazy and also no-parametric, and in this technique the results are compiled for the new systems from the database that includes separated data points that are clustered. It is profound and easy to use technique because of its simplicity, despite this; it can overkill the performance of other complex models especially in forecasting the economics situations and its related demographics and data compression etc. In the recommendation system settings the KNN collaborative filtering algorithm is used to crawl through the whole space of user items to determine the neighbors of the user. In this regard, it is intricate and thorny to provide the recommendation in real-time due to its unpleasant approach towards consumption of time while swarming through the large number of users.

II. REALTED WORK

A. Literature Review

To overcome the issue of Cold User (cold start) in existing techniques e.g. CF many researches proposed their models, architecture and frameworks. Before conducting this study, around 20 studies has been gone through that identified and addressed the same problem while conducting their work. These studies discussed the problem and also proposed their solution to overcome cold start problem. In this study the solution produced by the researchers are reviewed in detail and

sectioned according to their nature. The review of the solutions is carried out based on the categories like, cross domain and social network data system, implementation of association rules on the user profile, similarities in ratings and demographics, behavior and historical interest of the users, deep learning strategies, ICHM, CBCF, OWA and CART.

In 2015, another research addressed the cold start problem and generating recommendation for cross-selling items and products on large e-commerce websites like Amazon and eBay[5]. In this study it is mentioned that the cold start problem hinders the functioning of the recommender system when there is insufficient information about the user or inadequate ratings of the particular item. In this study, knowledge based links among the large e-commerce platforms and the sharing and transfer of knowledge about the user and the item among the domain can mitigate the cold start problem. Cross domain recommender system addressed the cold start problem by sharing and transferring information about the item or user from other similar platforms. This, solution has been implemented and now can be seen in many applications of renowned tourism organizations. Suggesting the tourism places from the picture uploaded and tagged by other tourists on the social media networking websites about the places which are not explored by the particular user is the effective application of the cross-domain recommender system [6] [7, 8].

In this regard, researchers developed an application in which cross-domain knowledge sharing technique is used to suggest the travel destination based on the tourists travel interests and suggests likewise destinations based on the geo-tagged pictures uploaded on the social media with the prediction of the weather[8]. In this study it is mentioned that the problem arises when a new tourist is going on a trip. The recommendations for the new tourists from the system are then generated by the other tourist trip planned for the last location. While matching the interests for the particular user from the likewise social media profiles of other users some of the irrelevant and non-related generation of recommendation occurs which infers in incorrect results thus reducing the efficiency of the whole recommendation system.

Fig. 1 shows the cross domain recommendation model in which recommendation for the new users are generated on the bases of their behavior and from the users of the same group matching their interests from the social media networking profiles. So the knowledge about the user behavior, their likes and dislikes is shared with the targeted domain.
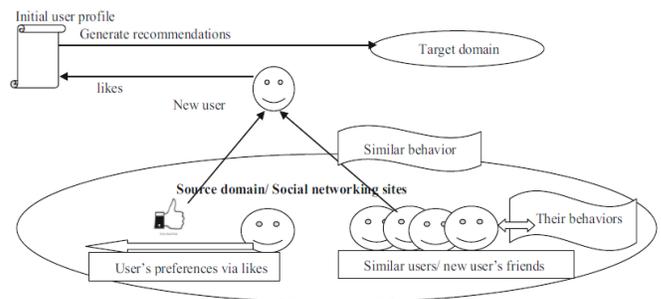

Fig. 1. Cross Domain Recommendation Model.

Shapira, Rokach, and Freilikhman [9] in their research discussed the cold start problem and suggested a solution for this problem. It is suggested that recommendation system can collect data from the Facebook profile of the user and filter out the data related to the particular domain. If there is insufficient information about the interests of the user related to the domain then it may extract the information from the behavior of the user and the reactions posted on the friends profile by that user. This study suggested the cross-domain learning method that is k-nearest neighbor which is source a source aggregation method for collecting the information about the new user in the specific domain. The authors concluded that this aggregation method based k-NN shows significant improvement in the recommended results if the users reacted to the content is not much spaced.

Thus, infers that the cross-domain model functionality is based on the data set density. For resolving the cold start problem the data set density is the significant variable.

Sobhanam concluded in their research that the cold start issue can be addressed by the implementation of association and clustering [10]. The association rules are implemented on the profile of the new user to generate a new enriched profile, also the implementation of the frequency pattern taxonomy based profile aided in the generation of these profiles of the user. This technique can generate the top-N recommendations for the new user.

Park addressed the same issue but in different domain [11]. This is relatively old study but addressed the cold start as it is based on the filterbots algorithm. This study discussed that the naïve filterbot algorithm is used by which the system can be infused with bots or hypothetical user. This algorithm can collect the information about the rating of the item (here item can also be a user). This model extracts the average rating for the item or user that is based on the attributes or association of the specific user or item. It also calculated the similarities between users using demographics of the users and determining its average rate. The infused bots and pseudo users in the user matrix as another user or actual rating of the item are treated by the system by applying algorithms of collaborative filtering for the generation of recommendations. Item based and user based algorithms are used to predict the interests of the user and to determine interest similarities among the users Pearson correlation is used which is given as:

$$S(a,b) = \frac{\sum_{i \in I_a \cap I_b}(r_{a,i}-\overline{r_a}) \cdot (r_{b,i}-\overline{r_b})}{\sqrt{\sum_i(r_{a,i}-\overline{r_a})^2} \cdot \sqrt{\sum_i(r_{b,i}-\overline{r_b})^2}}$$

Where:

S = similarity

a, b = users

$(r_{a,i} - r_a) \cdot (r_{b,i} - r_b)$ = difference between the item's rating by user a, b and average rating

$I_a \cap I_b$ = the set of items rated by the user a,b

In recommending the items to the new users based on the demographic similarities and user interest similarities, there is another approach other than the naïve filterbot model that is

Triadic aspect model. In the subsequent section some models are discussed with details as part of literature review.

Lam and his team in their research proposed a triadic aspect model which generates the prediction ratings of the item using information of the similar user as its input [12]. It extracts the statistical demographics like gender, age, work and other related information of the user. This model was suggested by the Hoffman to analysis the two-mode and co-occurrence data that has implementations in machine learning, retrieval of information and its filtration. This model predicts the interest of the users based on the triadic aspects includes age, gender and work, through which the interests of the existing users are extracted and generate recommendations for the new users with similar demographics and triadic features. This study concluded that the results are satisfactory when implemented to about 280 different types of new users but where the data set is relatively larger the results produced are not satisfactory and recommendations generated for the new user to address the cold start problem are not precise.

In a research paper it was narrated that predicting users based on their interests and behavior for generating recommendations related to them is like a web intelligence system [13]. The cold start problem discussed and addressed in their studies by suggesting implementation of clustering techniques to framework of item-based collaborative filtering. This research also suggests the integration of the information content into the collaborative filtering. It infers the hybrid approach towards the solution of the cold start problem that is item-Based Clustering Hybrid Method. The features of the content information are clustered through which the preferences of the user rating are complemented. A statistical approach is suggested in the study that linearly combined the likeness between the user ratings. It uses the cosine measures for the rating of clustered content and Pearson measures for the ratings of the users. It is concluded after experimental results that the sparsity in the data is effectively addressed and also significance improvement is noted in the recommendation [14]. It was not the only hybrid model used previously but number of researches suggested hybrid model as one such model was introduced in which collaborative filtering, content based and demographic based models were infused to address the cold start problem [15].

Meng used collaborative filtering method as a different approach to increase the performance of cold start problem [16]. The extortion of the user's interest is achieved by walking through the history interests and cognitive similarity among the alike users to implement the social sub community division through well-known similarity measurement of Pearson and clustering approach based on K-means. The recommendations were generated by building the CART based upon user's static information & distinct group of alike users.

Sakarkar and Deshpande used a dissimilar and unique approach to collect the information about the user. In this study [17] the past educational data of the user was subjected to the k-means clustering technique. It was a somehow like triadic model, but it uses three different features as an input of the recommendation model. Despite educational data, current professional experience and information of the parents are used

as an input. However, the classification of the new user is done by the k-nearest neighbor model based on his attendance in the computer based assessment. IMSAA online real dataset was used to carry out the experiment along with movie lens 100k dataset. This technique however, directly addressed the cold start for new user effectively.

Jazayeriy and his team in their study used the approach to effectively used the rating of the items to mitigate the cold start issue for new users [18]. This technique was based on the existing category of the item; it determines the average ranking of the each item in the category and judge against the user's average rating to improve the ranking of the recommended items. It shows an acceptable improvement while tested on the movie lens 100KB, 1MBand 10MB datasets.

B. Existing Recommender Algorithms

The cold start problem is addressed by various researchers in multiple ways however, it is established by now that the mentioned problem have its solution veiled in addressing the effectiveness and efficiency of the algorithms used to determine the input for the recommender model. This lead the study to infer that there are two most common proposed models; collaborative filtering and content based methods. The earlier one according to this study refers that the new user will receive the recommendations based on the likeliness of the item from other similar users with alike preferences. However, it also pointed that the features based analysis of the users is not mandatory. While the later solution according to this study refers that items are analyzed by the said methods (content based) or take in account the features of the users to extract the items from the internet that matches the interests of the new user. However, this study ponders on the distinctive situation of the cold start problem subjected to the condition that new user does not show up with any of the previous ratings. In such situation, both collaborative filtering and content based methods will function to a limited accuracy performing less effectively to address the issue. This way the recommendation will not be generated for the cold user, hence affecting the efficiency of the algorithms. This will elevate the quantity of objects present in systems than amount of objects that have been rated till then by the user especially through thin data availability and this intensify the contention of the cold start problem [19].

The above situation gives rise to two issues; cold start user and cold start item. The earlier issue arises due to insufficient activity turned up from the new user regarding item ratings, this in turns become difficult for the recommended system to find similarity among users. The later mentioned issue arises due to lack of ratings on items from other users, therefore less likely to be picked up by the recommender system.

CF approach will not be proposed keeping the scale of this research in accordance as this technique uses models of deep and machine learning algorithms to predict the rating of the item which were not rated by the users up to adequate level. There number of collaborative filtering algorithms that are based on the Bayesian Networks, semantic models and clustering models. Markov decision process, singular value decomposition and multiple multiplicative factor based models are the examples of model based collaborative filtering [20].

Collaborative filtering accomplished through such algorithms develops the model of the rating of the user at first for providing the recommendation of the item. Probabilistic approach is used for algorithms in this group for visualizing the process of collaborative filtering. In such situations the visualization or imagination of the collaborative filtering process behaves same as if the estimated value of prediction for user is computed based on the rating given by the user on other items.

However, the scope of the study requires that memory based model must be used instead as the contribution of the study is to find the effective way of determining the similarity between the users and among items. The memory based collaborative filtering is used to determine the list of items for the subjected users that is based on the similar behavior among the users. Memory based collaborative filtering can be discussed from both item based and user based perspectives. However, according to the model proposed in this study user-based collaborative filtering will be discussed in the subsequent section.

C. User based Approach

The user based collaborative filtering is adopted to determine the rating of an item received from multiple users, for this purpose the similarity among these users is calculated. Henceforth, the rating for the same item from the subject user can be predicted with the help of computation. As discussed earlier, Pearson correlation coefficient for subjected user will produce the association and connection among user "u" and "v" illustrated in eq 1 as:

$$\text{Sim }(u,v) = \frac{\sum_{i \in I}(r_{u,i} - \overline{r_u})(r_{v,i} - \overline{r_v})}{\sqrt{\sum_{i \in I}(r_{u,i} - \overline{r_u})^2}\sqrt{\sum_{i \in I}(r_{v,i} - \overline{r_v})^2}} \tag{1}$$

Where;

$i \epsilon I$ and I = over all co-rated items from both u, v users

$r_{u,i}$ = rating of item "i" by the user "u"

$\overline{r}_u, \overline{r}_v$ = mean rating of the user "r" and "v"

D. Prediction Computation

The prediction for the subjected user for a specific item can be obtained by the formula given below in eq 2:

$$\text{Predict }(a,i) = \overline{r_a} + \frac{\sum_{u \in U} sim(a,u).(r_{u,i} - \overline{r_u})}{\sum_{u \in U}|Sim\ (a,u)|} \tag{2}$$

The above computation illustrated that, the summation of similarity of the user "a" and "u" and weights of the rating normalized by the sum of the weights on that specific item "i", then summing up this with the mean rating value of the active user "a" can produce the prediction of recommending that item "i" to the active user "a". Whereas, user "u" belongs to the neighbor user "U" in terms of high similarity index while (a,u) represents the range or set of the users with similar rating weights to which the specified user (a) belongs.

E. Fuzzy Clustering and C-means Algorithms

To determine the prediction of rating of an item for the specified user, computation of similarity of rating from the other similar users is achieved using statistical techniques. For

this purpose, this study proposes the clustering technique, in which the grouping of similar objects concealed in multidimensional and multivariate datasets and the partitioning of unordered items is done [21]. Using this technique, the scattered and uncategorized data can be gathered into clusters with similar type of items. This statistical approach is widely known in the machine learning enhancement, thus this study emphasize on the algorithm and technique of clustering in order to analyze the adjacency and likeliness of the users. For this reason, the clustering technique is used in the present study is mainly encompass Fuzzy algorithm. The fuzzy clustering is also known as soft clustering, unlike hard clustering the data points in the fuzzy clustering are not strictly belongs to a distinctive cluster but they can belong to more than one cluster at a time [22]. This is main approach of the study for using soft clustering, because the rating of an item may not be only represents all the similar users, but the item may be rated by the unique user or user from different cluster sometimes. The fuzzy clustering becomes very significant when these clusters do not have distinguishable boundary, and have overlapped behavior with neighboring clusters. Using this technique, proposed recommender system will not fall short when the exact similarity within the cluster is not available especially when the membership function is concerned. To achieve the fuzzy clustering the Fuzzy C-means algorithm is used which is known for identifying the fuzzy cluster with minimum cost function [23].

F. Clustering with Fuzzy C-Mean

Fuzzy C-mean algorithm is very similar to K-means algorithm and is one of the most widely used fuzzy clustering algorithms. This algorithm chooses number of clusters and not only one. In this algorithm each data point is randomly assigned with coefficients for being in the respective clusters. This algorithm is iterative until converged but the change between the coefficients is not more the given threshold i.e. denoted by ϵ. According to the scope of this study algorithm determine the clusters for the data point sets "x". This algorithm tries to partition the limited number of collection of n elements i.e. $X= \{x1……..xn\}$ if the criteria is given for being into the collection of c fuzzy cluster. In this perspective of given finite data set a partition matrix is returned with a list of c cluster centers i.e. $C= \{c1…….cc\}$ and illustrated as:

$$\underset{c}{\mathrm{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^{m} \left\| X_i - C_j \right\|^2$$

Where:

$$w_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| X_i - C_j \right\|}{\left\| X_i - C_k \right\|} \right)^{\frac{2}{m-1}}}$$

However, for a set of data point $X_j \in R_d$, where j=1……N with minimum cost function to find the partition where data point belongs is given as in eq 3:

$$j(U, M) = \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{i,j}^{m} D_{i,j} \qquad (3)$$

Where;

$\mu = [\mu_{i,j}]$, i denotes cluster, j denotes its object and μ overall represents the value of fuzzy membership in given cluster of given object and m belongs to the range of $[1, \infty]$.

However, the value of the distance between xj and mi is denoted by $D_{i,j} = D(x_j, m_i)$.

The algorithm in the continuity of above formulation can be developed as:

Step 1 Value of the m, c and ϵ will be defined along with the assigning step value t = 0 and initializing the mean matrix.

Step 2 When t= 0 it computes the membership matrix μ, else update when t<0 using the below formulation:

$$\mu_{i,j}^{(t+1)} = \frac{1}{\sum_{t=1}^{c} \left(\frac{D_{tj}}{D_{i,j}} \right)^{1/(1-m)}}$$

If i=1,….,c and j=1,….,N.

Step 3 From the preceding step, $\mu_{i,j}$ will take new values and the mean matrix C will be updated accordingly using:

$$C_i^{(t+1)} = \frac{\sum_{j=1}^{N} \left(u_{i,j}^{(t+1)} \right)^m x_j}{\sum_{j=1}^{N} \left(u_{i,j}^{(t+1)} \right)^m}$$

Step 4 Iteration of step 2 and 3 is required until the function converges and minimum cost value is achieved i.e. difference between the change in mean matrix is smaller than the small number ϵ.".

III. PROPOSED MODEL

Two algorithms are proposed as the framework of proposed model. One algorithm is for the training of the recommender system, where movie lens data set is presented to the algorithm as input and expected output of the recommender system will be fuzzy user-based measurement of similarity and briefs as:

START

Step 1 Access Movie Lens website to Load data set, to construct two matrices

Step 2 Construction of rating matrix of user-movie

Step 3 User based similarity matrix is produced from user-movie rating matrix i.e. preceding step, by implementing Pearson similarity measure.

Step 4 Truthfulness matrix of user is constructed

Step 4.1   Computation of activity of user u (count (Mu)= User_activity (u)

Step 4.2   Probity of user computation

$$\text{User\_Propity (u)} = \sqrt{\frac{\sum_{j \in M} (R_{u,j} - \overline{R_u})}{count (M_u)}}$$

Where;

R is rating

U is user

J is movie

Ru is average rating for the user (u)

Step 4.3   Computation of the sore of user's neighbor

MAX AVG (ui) = User_friend score (u) where i=1,…..,K(no. of nearest user)

Step 5 Similar users are clustered keeping degree of membership in account based on the truthfulness matrix after implementation of fuzzy c –means clustering. Where range of value of membership is [0,1]

Step 6 The fuzzy based matrix which infers the product of two similarity matrices (i,j) then produced by computing the proposed formula i.e. user-based similarity times fuzzy truthfulness similarity measure illustrated as:

Similarity (i,j)fuzzy based= Similarity user based x C (combination Coefficient) + similarity fuzzy truthfulness x (1-C)

Combining and taking the product of two matrices i.e. user based similarity matrix and similarity of fuzzy truthfulness matrix results in the fuzzy based similarity matrix as illustrated above.

END

The results of the first algorithm infers the fuzzy based similarity matrix, whereas, the second algorithm is used for the testing phase of the recommender system in which the new user with few ratings will be taken as input. This algorithm will produce the recommended items for the new user.

START

Step 1 selecting cold start users from the dataset

Step2 information for the new user's truthfulness is computed

Step 3 implementation of Pearson correlation coefficient on the new user's ratings matrix

Step 4 according to the new users truthfulness matrix clustering

Step 5 proposed similarity formula is computed i.e. $j(U,M) = \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{i,j}^{m} D_{i,j}$

Step 6 Implementation of prediction similarity formula (from prediction computation section)

Step 7 Fetch results of clustering of truthfulness matrix of new user from Step 4 (highest estimated ratings) and recommend to the new users.

END

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

in the title or heads unless they are unavoidable.

IV. Discussion and Conclusion

The model proposed in this study is tested through experiment by providing Movie Lens data to the developed algorithms. This data was consisting of more than 1600 movies and their corresponding 100 thousand ratings from more than 900 users, is used to extract the ratings form the user and process it further after analysis. Cold start users for this model are selected randomly from the data those who have less ratings. Scrutinize intimately Table I, it is translucent that there lies similarity between the ratings of movies from different users. However, some users have rated very few movies from first 10 movies and some of them did not rated at all. The data for the rating of the first 10 movies from 10 users is included in tabular form in Table I.

As an instance, user 3, 4 & 5 did not rate any of the movies, however user 9 rated only one movie. They have no common ratings, and also user 3, 4 and 5 did not rate first 10 movies. On the other hand user 8 and 9 rated only movie 7. For the purpose of computation and analysis user 4 and user 9 will be considered. In Table III, user 9 rated 12 movies and user 4 rated 13 movies. On measuring the similarity matrix based on user data the similarity between the users will be computed and is shown in the Table II.

It is clear from the user based similarity matrix that the value of similarity between user 4 and user 9 is zero. According to the algorithm 1 the processing is achieved offline which will definitely effective in reduction of recommendation time. The model proposed in this study is produced the user-movie matrix of rating as depicted in Table I and Table II is constructed when the Pearson correlation similarity formula is applied on the user-movie matrix.

However, the computation results of user activity and user probity are shown in Table III which illustrated the whole data set, which is obvious demonstration of the truthfulness of each user and their behavior.

In addition, fuzzy c-means clustering plays important role in the performance of prediction. According to the demand of the study the construction of fuzzy matrix is vital because producing clusters of truthfulness of the users to know about the belonging of the user which is illustrated in Table IV.

These values are the membership values of the user belong to the respective cluster and to be used to compute the similarity between users using the formula developed in the equation 3. After this computation the sparsity problem is mitigated, hence improve the prediction accuracy especially with cold start problem. The depicted two clusters are computed with fuzzy c-means which illustrates that, for instance, user 1 have 32% membership of cluster 1 and 68%membership of cluster 2. The accuracy in prediction cannot be achieved if it is being done only on the bases of user-based similarity matrix in user-based collaborative filtering especially for the users who does not bear common ratings for the same item. As for instance, user 4 is neither share ratings as of 6, nor as of 9, hence similarity value is zero. However, if the model proposed in this study is used, then the similarity and prediction can be achieved for the new users using the upgraded equation 3 according to which user truthfulness plays

the key role to successfully recommend the item to the cold user. From results and explanation of the suggested model it becomes translucent that mitigation of sparsity and cold user problem are well addresses through the vigorous fuzzy user-based collaborative filtering is effusive with the proposed framework.

TABLE. I.   USER-MOVIE RATING MATRIX

USERS	MOVIES									
	1	2	3	4	5	6	7	8	9	10
User1	5	3	4	3	3	5	4	1	5	3
User2	4	0	0	0	0	0	0	0	0	2
User3	0	0	0	0	0	0	0	0	0	0
User4	0	0	0	0	0	0	0	0	0	0
User5	0	0	0	0	0	0	0	0	0	0
User6	4	0	0	0	0	0	2	4	4	0
User7	0	0	0	5	0	0	5	5	5	4
User8	0	0	0	0	0	0	3	0	0	0
User9	0	0	0	0	0	0	4	0	0	0
User10	4	0	0	4	0	0	0	0	4	0

TABLE. II.   USER-USER SIMILARITY MATRIX

User	MOVIES									
	1	2	3	4	5	6	7	8	9	10
1	1.0000	0.9545	0.8555	0.9318	0.9285	0.9527	0.9401	0.9754	0.9690	0.9677
2	0.9545	1.0000	0.9522	0.9918	0.9829	0.9565	0.9624	0.9664	0.8907	0.9770
3	0.8555	0.9522	1.0000	0.9484	1.0000	0.8808	0.8721	0.8785	0.0000	0.9214
4	0.9318	0.9918	0.9484	1.0000	1.0000	0.0000	0.9058	0.9816	0.0000	1.0000
5	0.9285	0.9829	1.0000	1.0000	1.0000	0.9355	0.9036	0.9537	0.8807	0.9340
6	0.9527	0.9565	0.8808	0.0000	0.9355	1.0000	0.9579	0.9885	0.9583	0.9796
7	0.9401	0.9624	0.8721	0.9058	0.9036	0.9579	1.0000	0.9645	0.9337	0.9772
8	0.9754	0.9664	0.8785	0.9816	0.9537	0.9885	0.9645	1.0000	1.0000	0.9839
9	0.9690	0.8907	0.0000	0.0000	08807	0.9583	0.9337	1.0000	1.0000	0.9931
10	0.9677	0.9770	0.9214	1.0000	09340	0.9796	0.9772	0.9839	0.9931	1.0000

TABLE. III.   USER TRUTHFULNESS MATRIX

	User_activity	User_probity	User_Friend_score
User1	272	1.47	4
User2	62	1.75	4
User3	41	1.54	4
User4	13	0.33	3
User5	274	2.48	4
User6	200	1.32	5
User7	390	1.76	3
User8	52	02.44	3
User9	12	0.22	2
User10	630	1.37	3

TABLE. IV.   CLUSTERING TRUTHFULNESS INFORMATION MEMBERSHIP VALUE

Clusters	USERS									
	1	2	3	4	5	6	7	8	9	10
Cluster1	0.32	0.0094	0.0127	0.0028	0.0404	0.113	0.891	0.0107	0.0288	0.0558
Cluster2	0.68	0.9906	0.9873	0.972	0.9596	0.887	0.109	0.9893	0.9712	0.9442

REFERENCES

[1] Samarinas, C. and S. Zafeiriou, Personalized high quality news recommendations using word embeddings and text classification models. 2019, EasyChair.

[2] Zhao, J., H. Wang, and H. Zhang, A Regression-Based Collaborative Filtering Recommendation Approach to Time-Stepping Multi-Solver Co-Simulation. IEEE Access, 2019. 7: p. 22790-22806.

[3] Bharti, R. and D. Gupta, Recommending Top N Movies Using Content-Based Filtering and Collaborative Filtering with Hadoop and Hive Framework, in Recent Developments in Machine Learning and Data Analytics. 2019, Springer. p. 109-118.

[4] Abdelwahab, A., et al. Collaborative filtering based on an iterative prediction method to alleviate the sparsity problem. in Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services. 2009. ACM.

[5] Cantador, I., et al., Cross-domain recommender systems, in Recommender systems handbook. 2015, Springer. p. 919-959.

[6] Arain, Q.A., et al., Intelligent travel information platform based on location base services to predict user travel behavior from user-generated GPS traces. International Journal of Computers and Applications, 2017. 39(3): p. 155-168.

[7] Memon, I., et al., Travel recommendation using geo-tagged photos in social media for tourist. Wireless Personal Communications, 2015. 80(4): p. 1347-1362.

[8] Memon, M.H., et al., GEO matching regions: multiple regions of interests using content based image retrieval based on relative locations. Multimedia Tools and Applications, 2017. 76(14): p. 15377-15411.

[9] Shapira, B., L. Rokach, and S. Freilikhman, Facebook single and cross domain data for recommendation systems. User Modeling and User-Adapted Interaction, 2013. 23(2-3): p. 211-247.

[10] Sobhanam, H. and A. Mariappan. Addressing cold start problem in recommender systems using association rules and clustering technique. in 2013 International Conference on Computer Communication and Informatics. 2013. IEEE.

[11] Park, S.-T., et al. Naïve filterbots for robust cold-start recommendations. in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. 2006. ACM.

[12] Lam, X.N., et al. Addressing cold-start problem in recommendation systems. in Proceedings of the 2nd international conference on Ubiquitous information management and communication. 2008. ACM.

[13] Li, Q. and B.M. Kim. Clustering approach for hybrid recommender system. in Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003). 2003. IEEE.

[14] Wu, Z., Y. Chen, and T. Li. Personalized recommendation based on the improved similarity and fuzzy clustering. in 2014 International Conference on Information Science, Electronics and Electrical Engineering. 2014. IEEE.

[15] Basiri, J., et al. Alleviating the cold-start problem of recommender systems using a new hybrid approach. in 2010 5th International Symposium on Telecommunications. 2010. IEEE.

[16] Meng, C., et al. A method to solve cold-start problem in recommendation system based on social network sub-community and ontology decision model. in 3rd International Conference on Multimedia Technology (ICMT-13). 2013. Atlantis Press.

[17] Sakarkar, G. and S. Deshpande, Clustering based approach to overcome cold start problem in intelligent e-learning system. International journal of latest trends in engineering and technology (IJLTET), 2016. 7(1).

[18] Jazayeriy, H., S. Mohammadi, and S. Shamshirband, A fast recommender system for cold user using categorized items. Mathematical and Computational Applications, 2018. 23(1): p. 1.

[19] Xu, Z. and Q. Fuqiang. Collaborative filtering recommendation model based on user's credibility clustering. in 2014 13th International Symposium on Distributed Computing and Applications to Business, Engineering and Science. 2014. IEEE.

[20] Su, X. and T.M. Khoshgoftaar, A survey of collaborative filtering techniques. Advances in artificial intelligence, 2009. 2009.

[21] Javed, K., M. Iqbal, and S. Mehak, Competitive Analytics from Social Media for Five Leading Sportswear Stores. INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS AND SOCIAL SCIENCES, 2019. 9(4).

[22] Kruse, R., C. Döring, and M.-J. Lesot, Fundamentals of fuzzy clustering. Advances in fuzzy clustering and its applications, 2007: p. 3-30.

[23] Tidke, B., R. Mehta, and D. Rana, A novel approach for high dimensional data clustering. International Journal of Engineering Science and Advanced Technology (IJESAT), 2012. 2(3).