An Effective Approach of Feature Selection for Recommender Systems using Fuzzy C Means clustering along with Ant Colony Optimization and Neural Networks

Hitesh Hasija
Senior Software Engineer
VISA Consolidated Support Services India (Private) Limited
Bangalore, India
mailto:hitoo.hasija@gmail.com

Abstract— **Ratings provided by user for a movie or TV program also depends on demographic information, day type, time, mood, and other factors of environment. Thus, context aware recommender systems are used for providing recommendations. As recommender systems suffers from cold start, sparse data and scalability problems. Therefore, artificial neural networks (ANN) are used to solve these problems. But, training of ANN again takes a very large time, due to high dimensional features. Hence, feature subset selection is done to solve this problem. But, it is a kind of combinatorial optimization problem, which could be solved by using Ant Colony Optimization (ACO). ACO with heuristic function as fuzzy values obtained after applying fuzzy c means clustering algorithm over movie lens data set provided optimum results. Back propagation algorithm has been used, for the training of neural network, only on those features which are obtained after applying modified ACO along with roulette wheel selection algorithm. Results are analyzed by providing ratings of those movies which are not provided in "movie lens data set" yet, as a solution to cold start problem. Finally, the accuracy of recommender systems is calculated by determining mean absolute error and comparison is provided with other previous approaches for different data sets.**

Keywords— *Feature subset selection problem, Ant colony optimization, artificial neural networks, Fuzzy c means clustering, back propagation algorithm, and Roulette wheel selection*

I. Introduction

Recommender Systems are used to provide recommendations for different available items. They are basically classified into 3 major categories [1]: -

1. Content based recommender systems,
2. Collaborative based recommender systems,
3. Hybrid based recommender systems.

In content based recommender systems, the items are recommended to the user based on past user ratings. The drawback with content based is, it can provide recommendations only for the similar content sources, and cannot use them for other content types. In collaborative recommender systems, the items are recommended to the user based on the similar liking of users. But, recommender system also suffers from 3 major problems [2]:-

a. Cold start problem:-systems require existing data on a user in large amount, to provide accurate recommendations.
b. Scalability problem:-a very large computation power is required by the systems, because millions of users and products are available.
c. Data Sparse problem:-the users do not provide proper ratings to all the available items.

In hybrid recommender systems, based on the mixture of above two methods the items are recommended to user. But recommender systems, also depends upon context into which user's provide recommendations. In [3], Context is defined as information used to characterize a person, object or places, which are relevant for interacting a user and its application. Context aware recommender systems are based on the theme of environment around the user while he/she was rating any movie or liking any TV program. To solve scalability problem [4], TV Programs could be best described based on three vectors and they are relaxing, emotional, and informative content into them. Artificial Neural networks (ANN) are used to provide recommendations with context aware recommender systems, in order to solve these problems. An ANN is defined by three basic parameters:-

1. Connections between different layers of neurons.
2. Methodology used for updating the weights of interconnection.
3. Function used for converting neuron's weighted input to output.

An ANN consists of inputs (synapses) which are multiplied by weights (the strength of a signal) and then applied to a mathematical function to compute the activation of a neuron. Another function computes the output of that neuron based on a certain threshold and these functions could be any linear or nonlinear activation functions. The weights could be any positive or negative value lying between -1 to +1. A positive weight value is used to increase the strength of input signal applied and a negative value is used to decrease the strength. Thus, we can get the desired output by adjusting the weights and this approach is known as learning or training the neural network. But, a real time neural network consists of hundreds of neurons in input and hidden layer, so it is quite difficult to adjust weights by hand. In order to achieve this, several algorithms have been proposed and one of them is back propagation algorithm. It uses supervised learning mechanism, i.e. for a given set of input values to be applied to input layer neurons, their corresponding output values to be obtained from output layer neurons is provided to us. The difference in target output and obtained output is known as error. The aim of this algorithm is to reduce the error by adjusting the weights accordingly. Thus, errors are propagated in backward direction to train the network.

The paper consists of 5 sections, section II consists of background details, section III describes methodology used to solve the problem, section IV covers results and comparison with previous approaches, and section V provides conclusion and future work.

II. BACKGROUND

Christakou and Stafylopatis had proposed a movie recommender system [5] where, in which the system acts as a classifier and determines whether a user would like the movie or not, by using artificial neural networks for predicting ratings of user. Kwon and Hong had proposed a solution for cold start problem by providing faster learning convergence method in [6]. In [7], it is shown that, with extensive machine learning algorithm, a single hidden layer feed forward neural network provides the same recommendation accuracy as that of back propagation algorithm. According to [8], values of weights could be assigned randomly between inputs to hidden layer, if the activation function used is infinitely differentiable. The examples of such activation functions are sigmoid, radial basis, sine, cosine, and many more non regular functions [9]. So, in the proposed approach we are using sigmoid function for the activation of hidden layer neurons and hyperbolic tangent function for the activation of output layer neurons.

[10] and [11] Suggests that, all the contextual information should be used to provide recommendations like mood, interests, demographic information, date, time and day, etc. Thus, no. of neurons in input layer should be equal to the features used. As the available movie lens data set are described by 24 dimensional features like action, adventure, crime and comedy, etc., so considering all features increases the time complexity of our neural network. Hence, for reducing the dimensions of feature vector, an approach based on principal component analysis had been proposed in [12]. It reduces the number of neurons required in input layer and hidden layer, leading to low training time. Feature selection is considered as one of the discrete optimization problem because the whole search space consists of 2^n possible subsets [13]. Feature selection is usually done in machine learning. Basically it is used for dimensionality reduction. The main objective of feature selection is to reduce down a very large set in our datasets to a small subset of features [14]. It has very wide applications like in data mining, text categorization, machine learning, etc. Selection of features should be done very carefully because it improves the efficiency of linear classifier used further and thus increases the performance of system with low computational complexity [15]. Out of so many methods used for feature selection, evolutionary algorithms based on population of data sets have shown a greater performance, because their future iterations depend upon past iterations. ACO algorithm is Meta heuristic in nature and it was proposed by M.Dorigo and its colleagues in 1990 [16]. ACO algorithm works on the principle of pheromone laying behaviour of ants [17].

Feature selection is categorized into filter and wrapper methods [19] and [20]. If the feature selection method does not include any learning algorithm, then it is filter based approach. It works on the principle of inter class separability criterion like k nearest neighbour approach. On the other hand, wrapper method uses the concept of learning algorithm. Wrapper method produces much better results as compared to filter method, but they are expensive to run on very large datasets [21]. A hybrid approach based on ACO for feature selection had been proposed in the forecaster [22]. In this paper we are going to use a hybrid approach, but it would depend upon ACO and fuzzy values obtained after applying fuzzy c means clustering algorithm on different movies with respect to contextual attributes as its heuristic function. In fuzzy c means clustering algorithm, the value of membership function is initially, assigned a random value from 0 to 1. This value of membership function is modified in further iterations based on the probability distribution of different items in different clusters. In different iterations, the main focus of algorithm is to minimize, the value of objective function [23], which is defined as follows with Equation 1:-

$$J_m(P) = \sum_{k=1}^{n} \sum_{i=1}^{c} [A_i(x_k)]^m \parallel x_k - v_i \parallel^2 \tag{1}$$

In this equation (1), outer loop runs n number of times, where n is the number of data points taken. Inner loop runs c number of times, where c is the number of clusters initially defined. In this case, c is equal to the number of contextual attributes taken for our data set. M is defined as fuzziness coefficient; it decides the measure of overlapping provided between different clusters. If the value of m is very large, then it means that a higher overlapping is provided between different clusters. But, on the other hand if the value is very small, then overlapping is not possible at all, and there would be no reason of using fuzzy c means clustering algorithm over there. The value of m could range from 1 to infinity. X_k is the value of data point to be considered, by subtracting its value from $V_{k,}$ where, $V_{k\ is}$ the value of center point of that cluster, which is considered in the inner loop. Lastly, A_i is the value of membership function, to be calculated or modified again and again in every loop, so that its fuzziness of data point with respect to a cluster could be determined. Its value is modified for all iterations, according to the formula: -

$$A_i^{(t+1)}(x_k)=[\sum_{j=1}^{c}(\frac{\| x_k - v_i^{(t)} \|^2}{\| x_k - v_j^{(t)} \|^2})^{\frac{1}{m-1}}]^{-1} \qquad (2)$$

The value of membership function is modified as per the rule given in Equation (2), and accordingly, the value of center for different clusters is also modified in all iterations, as data point's increases or losses the membership values with respect to different clusters. The formula used for center calculation is as follows, defined in Equation (3): -

$$v_i = \frac{\sum_{k=1}^{n}[A_i(x_k)]^m x_k}{\sum_{k=1}^{n}[A_i(x_k)]^m} \qquad (3)$$

Finally, in the last step, we have to determine the stopping condition. It is defined, in terms of some threshold value, which could be taken as 0.15 and its equation is defined as follows; -

$$| P^{(t+1)} - P^{(t)} | = \max | A_i^{(t+1)}(x_k) - A_i^{(t)}(x_k) | \qquad (4)$$

Hence, if | P (t+1) - P (t) |≤ε, then stop as described in Equation (4), otherwise repeat all the steps of membership function calculation and center calculation as mentioned above.

III. Methodology Used to Solve The Problem

A. Ant Colony Optimization

Ant colony optimization learns from its past experiences to determine the best trail for future iterations as referenced by [24]. ACO works on two basic parameters:

a) The pheromone updating rule, and
b) The heuristic information for converting our problem to maximization or minimization,

ACO with heuristic information provides better results as compared to without heuristic information. In this paper, Ant colony optimization is used to solve the feature selection problem. For applying ACO, firstly conversion of feature selection problem into graph form as described in [25] is required. Then, we have to define the pheromone values i.e. the weights of interconnection between different nodes. These weights are obtained based on the values of membership function of different features with respect to ratings provided by user for corresponding movie or TV program. These fuzzy values are obtained by applying fuzzy c means clustering algorithm as heuristic function with ACO. So, the actual problem of feature selection has been converted to maximization of these fuzzy values. Because, max value of membership function signifies that the rating of movie or TV program varies more with respect to given feature subset as compared to other feature subsets. While applying ACO, the paper uses the Fisher Yates shuffle algorithm to randomize the order of traversing nodes initially. Pheromones are chemicals placed by ants while traversing a particular path. As the ant reaches to the destination earlier, more the amount of pheromones would be laid down by it and more the number of ants travel through that path. If a path is not used, then corresponding pheromones will slowly evaporate over time. ACO implemented in this paper is in modified form along with roulette wheel selection algorithm as referenced by [26].

B. Merging of ACO with Fuzzy C Means Clustering

Fuzzy c means clustering algorithm is used to divide the given data sat of n items into different partitions or clusters which are already defined in number. Thus the number of clusters would be equivalent to the different contextual attributes. As initially, in the given table we are usually provided with the movies as the items with first column and genre or contextual attributes as first row. So, we have to start with a tabular matrix, drawn between movies and its contextual attributes. Corresponding to each movie, we are provided with the values of its contextual attributes that whether the movie consists of comedy, romance, adventure, action or not. Hence, these values of different contextual attributes act as data point in an n dimensional plane where, n is the number of contextual attributes. After applying fuzzy c means clustering algorithm over to it, the clusters get formed as per the contextual attributes defined. Now, these clusters, consists of different movies into them with some membership function defined for each item or movie. As fuzzy c means clustering algorithm, says that an item could also belongs to more than one cluster but with different values of membership function defined over to it. Hence, in this case also, one movie could also belong to more than one cluster, with different membership function values defined over to it. Finally, we get the result in the form of tabular matrix again, but between different movies or items and clusters obtained. From here, we could obtain the values of different contextual attributes with respect to different movies, in the form of membership function defined. Hence, these values of membership function are added to obtain the contribution of different contextual attributes in different movies, which is further used as the heuristic function in ACO.

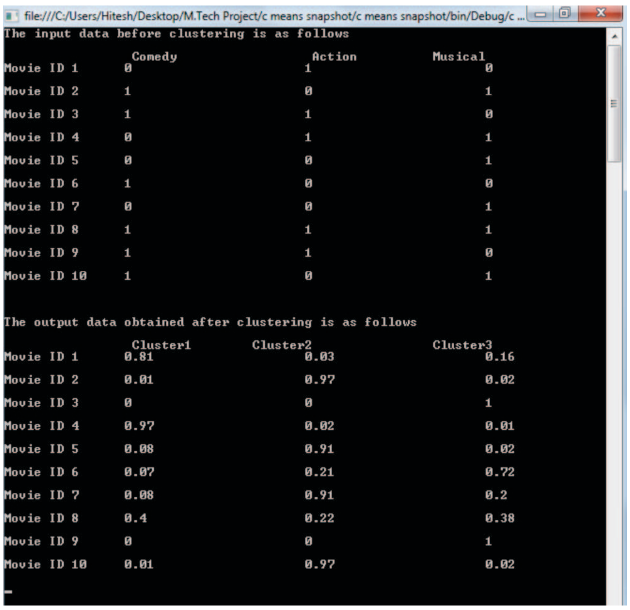
C. Merging of ACO, Fuzzy C Means with Back Propagation algorithm

After obtaining a feature subset by applying ACO, number of neurons in the input layer is determined. Suppose, the feature subset consists of 5 features, then 5 input neurons should be used. Predictions provided by user are not linear in nature, thus to predict nonlinear variations, the activation functions should be used accordingly. Sigmoid function is used very commonly to predict the ratings of users, as it is nonlinear and squeezing function. The sigmoid function is very close to 1 for large positive numbers, 0.5 at 0 and close to 0 for large negative numbers. Hence, this paper uses sigmoid function to provide activation of neurons at hidden layer and hyperbolic tangent function for activation of output layer neurons. In order to reduce the time complexity, 3 hidden layer neurons are used. The output layer will consists of 5 neurons corresponding to 5 different rating i.e. 1, 2, 3, 4, and 5. Neural network is trained by applying given set of input values provided in the data set. Initially random weights and biases are assigned to the neural network, lying from -1 to +1. In the first epoch, error is computed by comparing the values of output obtained and target output values to be achieved. Then, the error is propagated backwards as per the rule of back propagation and weights are updated accordingly. These epochs repeat themselves until the value of error becomes smaller than 0.01 or unless 1000 epochs gets completed. In the similar way, values of weights and biases are computed for different instances of data, but only those contextual features are considered, which lies in the subset obtained by applying modified ACO to it.

After training of neural network, the new sets of instances are used to calculate the accuracy of the neural network. If proper ratings would have been provided by the network same as that of target output to be obtained, then it is considered as properly classified otherwise misclassified. As, individual difference between the target output and actual output is computed, if the difference is less than 0.1 for all output neurons, then test sample has been considered as correctly classified otherwise misclassified. Accuracy for training data is computed by using the formula: -

$$Accuracy = \frac{number\_of\_samples\_correctly\_classified}{Total\_number\_of\_samples} * 100 \quad (5)$$

IV. RESULTS & COMPARISONS


file:///C:/Users/Hitesh/Desktop/M.Tech Project/c means snapshot/c means snapshot/bin/Debug/c ...

The input data before clustering is as follows

	Comedy	Action	Musical
Movie ID 1	0	1	0
Movie ID 2	1	0	1
Movie ID 3	1	1	0
Movie ID 4	0	1	1
Movie ID 5	0	0	1
Movie ID 6	1	0	0
Movie ID 7	0	0	1
Movie ID 8	1	1	1
Movie ID 9	1	1	0
Movie ID 10	1	0	1

The output data obtained after clustering is as follows

	Cluster1	Cluster2	Cluster3
Movie ID 1	0.81	0.03	0.16
Movie ID 2	0.01	0.97	0.02
Movie ID 3	0	0	1
Movie ID 4	0.97	0.02	0.01
Movie ID 5	0.08	0.91	0.02
Movie ID 6	0.07	0.21	0.72
Movie ID 7	0.08	0.91	0.2
Movie ID 8	0.4	0.22	0.38
Movie ID 9	0	0	1
Movie ID 10	0.01	0.97	0.02

Figure1. Snapshot as implementation of basic fuzzy c means clustering algorithm.

The given snapshot mentioned below in figure (1), is a simple example of fuzzy c means clustering algorithm, in which 3 contextual attributes are taken i.e. comedy, action and musical. Every row corresponds to one movie and represented by its Movie ID. Every movie has certain characteristics i.e. every movie is related with some genre which are associated with it. In movie lens data set, there are about 19 genres given which represents different characteristics of a movie. For example, as given in the above mentioned snapshot, Movie ID 1 consists of action, thus corresponding genre of that movie is 1 and rest are 0. A movie could also consist of different genres into it, as mentioned above Movie ID 3 consists of comedy as well as action into it. In the similar way Movie ID 8 consists of all three genres i.e. comedy, action as well as musical. Hence, all the three bits are 1. Now, after applying k means clustering algorithm, we get the following 3 clusters: -

I. Cluster 1:- Movie ID 1, Movie ID 4.
II. Cluster 2:- Movie ID 2, Movie ID 5, Movie ID 7, and Movie ID 10.
III. Cluster 3:- Movie ID 3, Movie ID 6, Movie ID 8, and Movie ID 9.

But, as we know that in k means clustering algorithm, a movie could either belong to a particular cluster or do not belongs at all. In order to represent those movies which, consist of some features of musical, some features of action and some features of comedy, fuzzy c means clustering algorithm is used. As we can see from the output that, Movie ID 1 consists of only action, thus its contribution in Cluster 1 is 0.81 and its contribution in all the other clusters is negligible i.e. 0.03 in cluster 2 and 0.16 in cluster 3. Similarly, Movie ID 8 consists of all the genres, therefore it belongs to cluster 1 with a membership function value of 0.40, it belongs to cluster 2 with a membership function of 0.22 and it belongs to cluster 3 with a membership function value of 0.38. After applying fuzzy c means clustering algorithm, we could determine the contribution of every genre or contextual attribute in different movies and this contribution would be acting as the value of heuristic function further in ACO. Optimum feature subsets obtained for some standard data sets are as follows: -

1. Movie lens data set: - year, genre1, timestamp, companion, weather.
2. "LDos CoMoDa" data set: - age, day type, mood, social, time.
3. "In Car Music" data set: - natural phenomenon, landscape, sleepiness, traffic conditions, weather.

In order to solve cold start problem existing with "movie lens dataset", as so many contextual attributes are associated with it, so taking a proper subset of them, obtained by applying ACO with fuzzy c means clustering as heuristic function and training our neural network with this approach provides us the following results:-

	33	1	10	35	18	19	48	45	36	24	2	6	9	13	17
	3	5	5	2	5	3	3	1	4	3	5	2	5	5	5
	5	3	5	3	1	2	3	4	2	4	5	5	5	5	5
	3	5	4	2	5	3	5	1	5	5	4	1	5	5	4
	4	5	5	1	2	1	4	1	2	4	4	2	3	4	4
	3	4	5	1	5	3	3	2	1	3	5	2	5	5	5
	4	4	4	3	1	2	4	1	1	4	4	2	4	4	4
	4	4	4	2	5	5	4	3	4	4	4	4	4	4	4
	4	4	5	5	5	5	4	2	1	5	5	2	4	4	5
	5	4	5	1	5	5	5	2	5	5	5	1	5	5	5
	4	4	4	4	1	2	4	2	5	5	4	1	4	4	4
	3	2	4	1	4	4	3	2	5	3	4	1	4	3	4
	3	1	3	5	1	3	4	4	5	5	3	4	4	3	4
	3	2	4	1	5	5	5	4	2	4	4	4	4	4	4
	1	1	2	2	2	2	1	3	5	1	3	1	4	3	4
	5	2	4	3	3	1	4	3	1	4	5	2	5	4	5

Table1. Solution to cold start problem for some movies in Movie lens data set

In this Table (1), the very first row with green background cells represents the User ID's and similarly very first column represents the movie ID's corresponding to "movie lens data set". The cells marked with white background are those entries which are already provided in the data set. But, cells marked with red background are those, which are not provided in our data sets, rather which has been determined with our neural networks, after training it with modified ACO as described before. Now, the difference obtained in recommending top 5 movies out of these 15, without using this approach and with using this approach are as follows: -

User ID's	Top 5 movies before applying algorithm	Top 5 movies after applying algorithm
33	9,5,12,3,8	9,8,15,2,3
1	1,3,4,5,6	1,3,4,9,5
10	1,2,4,5,8	9,1,2,5,8
35	8,12,10,2,6	8,12,10,2,6
18	13,1,3,5,11	7,8,9,13,1
13	9,3,8,15,10	1,5,9,2,8
48	9,3,15,8,10	9,3,8,15,10
45	2,7,12,13,3	2,12,13,7,3
9	3,15,10,12,14	1,2,5,9,8
2	15,11,1,2,7	2,9,1,5,8

Table2. Change in recommendations for some users of Movie lens data set

As mentioned in the given snapshot of figure (2), implementation for the calculation of mean absolute error with 4 clusters has been done. With this approach, as we are already aware that in fuzzy c means clustering algorithm, we have to define initially the number of clusters. The numbers of clusters are equivalent to the number of possible subsets of contextual features obtained. So, here the numbers of clusters

Figure2. Calculation of Mean absolute error with 4 clusters

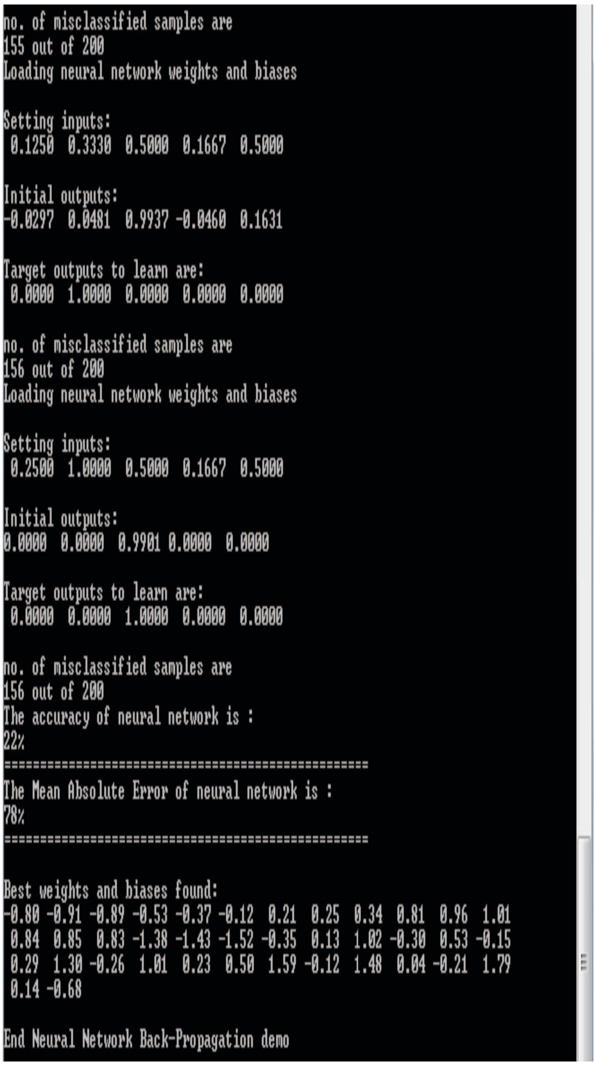


Figure2. Calculation of Mean absolute error with 4 clusters

are already defined to be 4, and accordingly the membership values of different movies are obtained. Finally, the determination of perfect feature subset is done by using ACO. Results are obtained by training the neural network on that perfect feature subset over 80% of our data set and then observing its behavior by testing it on rest 20% of our data set. As in the mentioned snapshot, the values of weights and biases have been calculated, along with accuracy and mean absolute error in our network. Total number of samples taken over here are, 1000 in number i.e. 1000 users of "movie lens" data set are used, out of which, 800 are used for training the neural network and rest 200 are for testing. Misclassified samples are obtained by comparing the obtained ratings from actual ratings. As 5 neurons in the output layer are used, thus if the value obtained as an output in any neurons is obtained to be more than 0.5, then that number of neuron is considered as provided rating. Suppose, if third neuron has a value of 0.97 and rest all have values less than 0.03, then rating 3 is obtained.

Mean absolute error has been calculated with 8 clusters i.e. initially while using fuzzy c means clustering, we have to define the number of clusters initially itself. So, if the numbers of contextual attributes are taken to be three, then total number of possible subsets calculated to be eight. Hence, it would be equivalent to the numbers of clusters. After defining number of clusters, membership function values of different movies are determined in all these clusters. Then, ACO is applied by using these values of membership function as heuristic function in ACO. A perfect feature subset out of those eight possible feature subsets is obtained as a result from ACO. Again, for those obtained feature subsets, neural network is trained and training as well as testing is done. Results are calculated by providing the ratings of 200 users out of 1000 users, and comparing actual ratings with the ratings obtained.

The experiment is performed on 16 clusters, and the results i.e. mean absolute error is obtained to 72%, while the accuracy which is just the reverse of mean absolute error is coming out to be 28%. Now, as defined in the approach, with those possible 16 subsets, the value of membership function is obtained and those values are further used as heuristic function in ACO. After applying ACO, a perfect feature subset is obtained, and with that feature subset the neural network is trained. After providing training to the neural network, testing is done on 200 users out of 1000 users and again the results are compared based on the ratings obtained and actual ratings.

Further, a graph has been drawn between mean absolute errors with different number of clusters to provide comparison with all the previous approaches [27]. Graph shown below, proves that efficiency of defined approach is better as compared to other previous approaches.

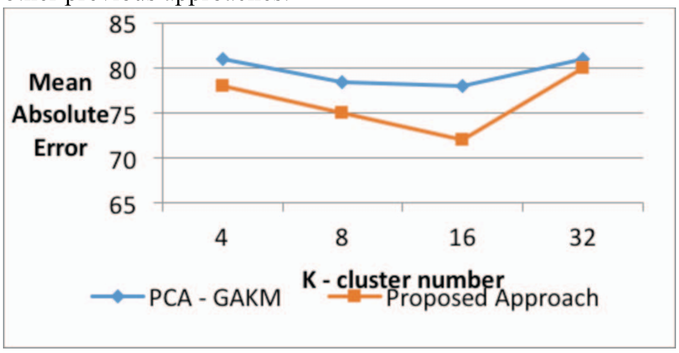


Figure3. Precision error on different cluster numbers

Finally, graph as described in Figure (3), between mean absolute errors calculated with two different approaches. This graph has been drawn between PCA-GAKM based approaches as described in [27] and the approach described in this thesis. As given in the snapshots pasted above in the experiments and results section, it has been clearly shown that mean absolute error with 4 numbers of clusters is 78%, with 8 numbers of clusters is 75%, and so on. Similarly, the graph also resembles the same thing. Thus, it proves that our approach is better than previous other approaches. Hence, this graph describes the precision error on different number of clusters.

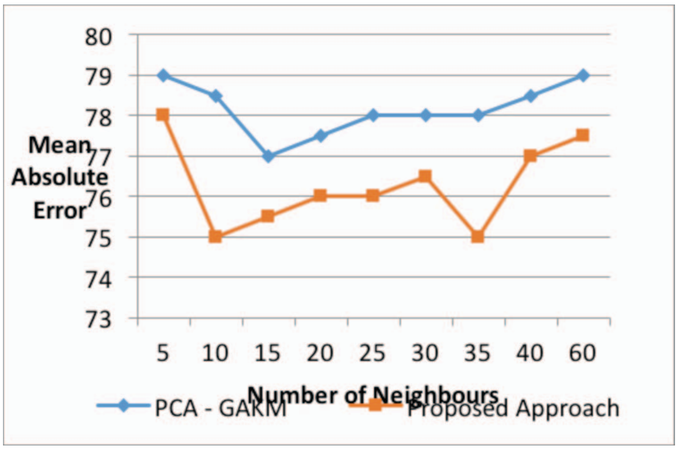


Figure4. Comparing accuracy with other clustering based approaches

As described in Figure (4), after providing a comparison between previously defined approach and our approach with respect to different number of clusters. Now, in the above mentioned graph, again the mean absolute error is calculated but this time the numbers of neighbors in a particular cluster are not kept fixed. The results are obtained with our approach, keeping number of neighbors as 5, 10, 15, 20 and so on up to 60, and the results are like for 5 neighbors the error has been reduced from 79% to 78%, for 10 neighbors from 78% to 75%, and so on. There is a downfall in graph two times, this downfall has been observed due to the activation functions used for hidden layer and output layer neurons. Sometimes this activation function correctly depicts the pattern of input and output and sometimes it may not be able to predict the variation of ratings with respect to users.

V. CONCLUSION & FUTURE WORK

In a nutshell, the recommendation system problems are addressed by selecting a feature subset using modified ACO with fuzzy c means values as heuristic function and neural network is trained with back propagation to obtain recommendations and accuracy has been compared to previous approaches. ACO with fuzzy c means clustering values is very efficient evolutionary algorithm especially for optimization problems like discrete combinatorial. The set of feature vectors obtained through ACO, have a great influence on the ratings provided by user as compared to other subset of features. The given approach has been applied to the standard movie lens dataset and the results are analyzed by comparing the mean absolute error of neural network. Before training of neural network, the input sample of datasets is normalized between 0 – 1, so that proper results could be obtained by keeping computational complexity very low, as number of iterations/epochs required with normalized input samples are less as compared to raw data inputs. As a result, mean absolute error with ACO as feature subset selection has been decreased compared to without ACO. As far as, the future work is concerned then ACO with other similarity measures as heuristic function could also be applied to content based and collaborative based recommender systems, along with back propagation.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin, Toward the next generation of recommender systems: a survey of state-of-art and possible extensions, IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.

[2] Sanghack Lee and Jihoon Yang and Sung-Yong Park, Discovery of Hidden Similarity on Collaborative Filtering to Overcome Sparsity Problem, Discovery Science, 2007.

[3] A. Dey, Understanding and using context, Personal and Ubiquitous Computing, vol. 5, no. 1, pp. 4-7, 2001.

[4] T. Isobe, M. Fujiwara, H. Kaneta, T. Morita, and N. Uratani, Development of TV reception system personalized with viewing habits, IEEE Trans. Consumer Electron., vol. 51, no. 2, pp. 665-674, May 2005.

[5] C. Christakou and A. Stafylopatis, A hybrid movie recommender system based on neural networks, Proc. 5th Intl. Conf. Intelligent Systems Design and Applications (ISDA'05), pp. 500-505, 2005.

[6] H. J. Kwon, and K. S. Hong, Personalized smart TV program recommender based on collaborative filtering and a novel similarity method, IEEE Trans. Consumer Electron., vol. 57, no. 3, pp. 1416-1423, August 2011.

[7] M. Krstić and M. Bjelica, Personalized TV program guide based on neural network, Proc. 11th Intl. Symp. Neural Network Applications in Electrical Engineering (NEUREL 2012), pp. 227-230, 2012.

[8] G.-B Huang, Q.-Y Zhu, C.-K Siew, Extreme learning machine: theory and applications, Neurocomputing, vol. 70, no. 1-3, pp. 489-501, December 2006.

[9] G.-B. Huang, H. A. Babri, Upper bounds on the number of hidden neurons in feed forward networks with arbitrary bounded nonlinear activation functions, IEEE Trans. Neural Networks, vol. 9, no. 1, pp. 224-229, January 1998.

[10] F. S. da Silva, L. G. P. Alves, and G. Bressan, Personal TV ware: An infrastructure to support the context-aware recommendation for personalized digital TV, Intl. J. Computer Theory and Engineering, vol. 4, no. 2, pp.131-136, April 2012.

[11] S. H. Hsu et al., AIMED – A personalized TV recommendation system in P. Cesar et al. (eds.), Interactive TV: a shared experience, Lect. Notes Computer Science, vol. 4471, pp. 166-174, 2007.

[12] D. Mladenic, Feature selection for dimensionality reduction, Berlin Springer Verlag, 2006, ch-5

[13] Marko Krstic and Milan Bjelica, Context aware personalized program guide based on Neural Network, IEEE, 2012.

[14] H.R.Kanan, K Faez, and M Hosseinzadeh, Face Recognition system using ant colony optimization based selected features, IEEE Symposium on Computational Intelligence in Security and Defense Applications, pp. 57-62, USA, 2007.

[15] W. Shang, H. Huang, H. Zhu, Y. Lin, Y. Qu and Z. Wang, A novel feature selection algorithm for text categorization, Expert systems with applications, vol. 33, no. 1, pp. 1-5, July 2007

[16] M. Dorigo and G.D.Caro, Ant colony optimization: A new Meta heuristic, congress on Evolutionary Computing, 1999.

[17] E, Bonabeau, M. Dorigo and G. Theraulaz, Swarm Intelligence: from natural to artificial systems Oxford University Press, New York, 1999.

[18] V. Maniezzo and A. Colorni, The ant system applied to the Quadratic Assignment Problem, IEEE Trans. On knowledge and data engineering, vol. 11, No. 5, pp., 769-778, September 1999.

[19] R.O.Duda and P.E.Hart , Pattern classification and scene analysis, John wiley and sons., Chichester, 1973.

[20] G.Forman, An extensive empirical study of feature selection metrics for text classification, Journal of machine learning research 3, pp. 1289-1305, 2003.

[21] R. Jensen, combining rough and fuzzy sets for feature selection, Ph.D. dissertation, School of information, Edinburgh University, 2005.

[22] C.K.Zhang and H.Hu, Feature selection using the hybrid of Ant Colony Optimization and mutual information for the forecaster , in Proc. 4th International Conference Machine Learning and Cybernetics, vol. 3, pp. 1728-1732, Aug 2005

[23] Lin Zhu, Fu-Lai Chung, Shitong Wang. "Generalized Fuzzy C-Means Clustering Algorithm With Improved Fuzzy Partitions" [J]. IEEE Transactions on Systerms,2009:39-3.

[24] Kai YAN, GangXUE, Shao-wen YAO, An Optimization Ant Colony Algorithm for Composition of Semantic Web Services; 2009 IEEE.

[25] Mehdi Hosseinzadeh Aghdam, Naseer Ghasem-Aghaee and Mohammad Ehsan Basiri, Application of ant colony optimization for feature selection in text categorization, IEEE, 2008.

[26] Hitesh Hasija, An efficient approach for web navigation using Ant Colony Optimization, IEEE International Conference on recent advances and Innovations in Engineering, IEEE, 2013.

[27] Zan Wang, Xue Yu, Nan Feng, Zhenhua Wang, "An improved collaborative movie recommendation system using computational intelligence", Journal of visual languages and computing 25 (2014) 667-675.