Breast Cancer Data Analysis Using Weighted Fuzzy Ant Colony Clustering

S Nithya
Department of Computer Science and Applications
K.S.R College of Arts and Science, Tiruchengode
Nammakkal(Dt), Tamilnadu, India
nithu_nivi@yahoo.co.in

R Manavalan
Asst. Prof and Head
Department of Computer Science and Applications
K.S.R College of Arts and Science, Tiruchengode
Nammakkal(Dt), Tamilnadu, India
manavalan_r@rediffmail.in

Abstract- **The performance of Data partitioning using machine learning techniques is calculated only with distance measures i.e. similarity between the transactions is carried out with the help of distance measurement algorithms such as Euclidian distance measure and cosine distance measure. These measures did not consider the global connectivity. The Distance with Connectivity (DWC) model is used to estimate distance between transactions with local consistency and global connectivity information. The Ant Colony Optimization (ACO) techniques are used for the data clustering process. ACO is integrated with DWC to find spherical shaped cluster. The global distance measure model called DWC is enhanced with fuzzy logic. The transaction weights are updated using fuzzification process. All the attribute weight values are updated with a fuzzy set weight value. The distance with connectivity model is tuned to estimate distance between the transactions using the fuzzy set values. The distance measure model efficiently handles the uneven transaction distributions. The ant colony-clustering algorithm is also improved with fuzzy logic. The similarity computations are carried out with fuzzy distance measurement models. All the fuzzified attribute values are updated with weight values. Un-even data distribution handling, accurate distance measure and cluster accuracy are the features of the proposed clustering algorithm.**

Keywords- Distance with connectivity, Ant Colony Optimization, Fuzzy Ant Colony Optimization, Breast Cancer Dataset.

I. INTRODUCTION

Clustering plays a significant role in the field of research. Cluster is a collection of objects which are “similar” among them and are “dissimilar” to the objects belonging to other clusters. To help us better in our path of knowing more about clustering none other than Ants play an important role for identifying the clusters [1].Ant colony optimization (ACO) is a population-based meta-heuristic that can be used to find approximate solutions to difficult optimization problems. The ant colony clustering algorithm imitates the intelligent behavior of ant and applies it to the solution of hard computational problems. This work was initiated by Deneubourg [2]. To apply ACO, the ant randomly moving and picking, dropping the object to achieve cluster [3, 4, 5, 6]. The optimization problem is finding the best path on a weighted graph. The artificial ants incrementally build solutions by moving on the graph. The solution construction process is stochastic and is biased by a pheromone model, that is, a set of parameters associated with graph components (either nodes or edges) whose values are modified at runtime by the ants. In traditional ways like Euclidean distance and cosine distance were used to find out the connectivity among data, but the limitation involved was it can be done only in local consistency and not in global connectivity.

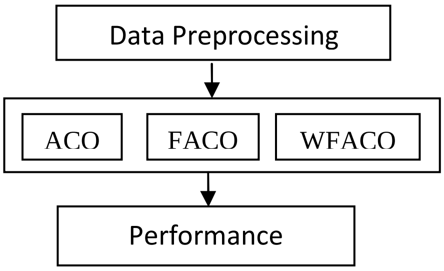

Figure1. Proposed Model

Therefore the appropriate solution was to go for DWC a distance calculating method [7]. One of the merits of DWC is to find inherent clustering characters of the data sets and suitable for data sets with uneven density distribution, comparing with the density-based methods. This study is needed to overcome the local consistency problem and to find the distance in global connectivity and further the accuracy has been increased by coalescing fuzzy logic with ACODWC. This paper is emphasized on clustering model to increase the accuracy and solving the local optimum problem. The proposed model is shown in Figure 1.

This paper is organized as follows: section 2 describes Literature survey, section 3 describes regarding Data Connectivity based distance estimation, section 4 depicts about Ant colony clustering algorithm based on DWC, Solution construction and Pheromone update rule, section 5 portrays on Fuzzy enabled clustering scheme, fuzzy logic concepts, Distance analysis, Fuzzification process, Fuzzy ACO and Weighted FACO, section 6 express in relation to Experimental analysis, and section 7 is end with wrapping up of the current work and upcoming enrichment.

II. LITERATURE SURVEY

Fuzzy ants and clustering projected a method Fuzzy c-means and hard c-means for reformulated the fuzzy partition validity metric and to clustering the data. Clustering web search results using fuzzy ants anticipated a method ACO simulated by means of fuzzy IF-THEN rules or fuzzy logic for clustering web search results have to be efficient and robust and cluster a dataset without using any kind of a priori information. Fuzzy controller designed by clustering-aided ant colony optimization proposed a method Ant colony optimization (ACO) algorithm for improving both the design efficiency of a fuzzy controllers and its performance. Fuzzy ant based clustering and fuzzy if–then rules for clustering the data with no initial partitioning and the number of clusters to be known in initial. An effective clustering algorithm with ant colony a method namely Sacc algorithm and Jaccard index for solving unsupervised clustering problem and to identify the optimal cluster number. A fuzzy-ACO method called Fuzzy rules and ACO for detecting breast cancer for solving problems in specific domain.

III. DATA CONNECTIVITY BASED DISTANCE ESTIMATION

DWC is used to find the distance in local compactness and global connectivity between the data. Establish the adjacency matrix of the data set. First, calculate the set $\phi_i = \{\phi_{ij} \in X, \ j = 1, \ \ldots, \ L\}$ of the data object $x_i \ (i = 1, \ \ldots, \ N)$ L (L> 0) nearest-neighbor [8], using the Euclidean distance formula. Then, link x_i and (i= 1, 2, …, N) L (L> 0) , and the link is undirected. First, nearest neighbor is computed using the Euclidean distance formula [8]. In this way, an undirected graph G = (X, V) and the adjacency matrix $R = [R_y]_{N \times N}$ of the data set, where $R_{ij} = \begin{cases} 1, & x_j \in \phi_i \ or \ x_i \in \phi_i \\ 0, & otherwise \end{cases}$ is constructed. V is the set of all links between data. Rs = [R_{ij}s] $_{N \times N}$, R_{ij}s is the number of s-steps reachability paths between x_i and x_j .

Definition 1: The connectivity between data object x_i and x_j is defined as: $Conn(x_i, \ x_j) = \sum_{s-1}^{step} conn^s(x_i, \ x_j)$

$$Conn^s(x_i, \ x_j) = \begin{cases} \frac{\log_L R_{ij}^{\ s}}{s-1}, \ if \ s > 1 \ \wedge R_{ij}^{\ s} > 1 \wedge i \neq j \\ 1, \ if \ s = 1 \wedge R_{ij}^{\ s} > 0 \ \ i \neq j \\ 0 \ , \ otherwise \end{cases}$$

The higher the connectivity between data point x_i and x_j is, the more reachability paths between x_i and x_j will have, which reflects the higher similarity between x_i and x_j , and the more close distance of x_i and x_j . Hence, DWC (x_i, x_j) ∞ 1/Conn (x_i, x_j). Furthermore, the definition of DWC (x_i, x_j) should reflect the local consistency at the same time, and then defined the DWC distance of data objects is defined as follows.

Definition 2: The DWC distance between data object x_i and x_j is defined as:

$$Conn^s(x_i, \ x_j) = \begin{cases} \frac{\log_L R_{ij}^{\ s}}{s-1}, if \ s > 1 \ \wedge R_{ij}^{\ s} > 1 \wedge i \neq j \ \text{ where} \\ 1, \ if \ s = 1 \wedge R_{ij}^{\ s} > 0 \ \ i \neq j \\ 0 \ , \ otherwise \end{cases} \qquad Dis(x_i, \ x_j) = \sqrt{\sum_{v=1}^{m} \left| x_{iv} - x_{jv} \right|^2}$$

It is the Euclidean distance between x_i and x_j, m denotes the number of the data object attributes, Max is a very large positive constant, M is a positive constant. If $Dis\ (x_i, \ x_j)$ is short and $Conn\ (x_i, \ x_j)$ is high,

$DWC\ (x_i,\ x_j)$ will be small. Then, the data object x$_i$ and x$_j$ will be clustered into the same cluster with a high probability. If $Dis\ (x_i,\ x_j)$ is short, but $Conn\ (x_i,\ x_j)$ is low, $DWC\ (x_i,\ x_j)$ will still be large. Then, object x$_i$ and x$_j$ will not be grouped in the same cluster. Obviously, DWC satisfies the following basic properties:

- DWC (x, y) ≥ 0, if and only if x = y, equality holds;
- DWC (x, y) = DWC (y, x).

DWC does not always satisfy the triangle inequality, so the definition of DWC is a generalized distance.

IV. Ant Colony Clustering Algorithm Based on DWC

The ant colony-clustering algorithm is improved with DWC [3]. Given X=$\{\mathbf{x}_1,\mathbf{x}_2,\ \ldots,\ \mathbf{x}_n\}$ a data set of N objects, and K (0 < K < N), form the number of clusters, clustering analysis organizes the N objects into K clusters, in order to minimize the clustering objective function F, where each object x$_i$ (i=1, 2, …, N) has m attributes, expressed as $\{x_{i1},\ x_{i2},\ \ldots,\ x_{im}\}$ [9]. The objective function is $Min\ F(w,\ C) = \sum_{j=1}^{K}\sum_{i=1}^{N} w_{ij}\ DWC(x_i,\ C_j)$, where C$_j$ is Centroid of Clusters (j=1, 2, …, k), and x$_i$ is an object subject to: $\sum_{j=1}^{K} w_{ij} = 1,\ i = 1, 2,\ \ldots,\ N$ and $\sum_{i=1}^{N} w_{ij} \geq 1,\ j = 1,\ 2,\ \ldots,\ K$. Here, w is an N-by-K weighting matrix, its elements: $W_{ij} = \begin{cases} 1, & if x_i \in cluster_j \\ 0, & if x_i \in cluster_j \end{cases}$

4.1 Solution construction

In ant colony algorithm, the ants construct solution (S). Ant, located at object x$_i$ (i = 1, 2 ,…, N), selects cluster C$_j$ (j = 1, 2, …, K) in probability P$_{ij}$ which is defined as $\mathrm{P}_{ij} = \frac{\tau_{ij}\,[\eta_{ij}]^{\beta}\,[path_{if}]^{\alpha}}{\sum_{k=1}^{K} \tau_{ik}\,[\eta_{ik}]^{\beta}\,[path_{ik}]^{\alpha}},\ j = 1, 2,\ \ldots,\ K$, where P$_{ij}$ is the probability distribution of object x$_i$ belonging to cluster C$_j$. $\eta_{ij} = \frac{1}{DWC\ (x_i,\ C_j)}$ delineate heuristic information value $DWC(x_i,\ C_j)$ which is the DWC distance between object x$_i$ and the center of cluster C$_j$. The heuristic factor β is indicating the relative importance of heuristic information. Path$_{ij}$ denotes the number of excellent ants [13], which construct good solutions and group xi into the cluster C$_j$. If path$_{ij}$ is very large, it can speculate that building good solution must group x$_i$ into the cluster C$_j$. The P$_{ij}$ reflects that if path$_{ij}$ is very large, then x$_i$ is grouped into the cluster C$_j$ with a high probability. By using this way a good solution has been developed rapidly.

4.2. Pheromone Update Rule

After each loop of the algorithm, i.e., when R (R ≥ 5) ants have completed a solution, then the solution is sorted according to the clustering object function value in ascending order. Then, it get S sorted = $\{s_1, s_2,\ \ldots,\ s_R\}$ where $s_q = \{c_{q1},\ c_{q2},\ \ldots,\ c_{qN}\}$, (q = 1, 2, …, R) of which use the top 20% better solutions $S\_best = s_q \in S\_sorted,\ 1\ \leq\ q\ \leq\ 20\%R \in Z$ to update the pheromone matrix using the formula $\tau_{ij}(t+1) = (1-\rho)\ \tau_{ij}\ (t) + \nabla\ \tau_{ij}(t)$ where $\Delta\tau_{ij}(t) = \left\{\begin{array}{l} \sum_{s_q S\_best} \frac{Q}{DWC\ (x_i, Cs_q\ j}, if c_q = cluster_j \\ 0, otherwise \end{array}\right\}$ $path_{ij} = \left\{\begin{array}{l} path_{i\ j}+1,\ is\ q \in S\_best \wedge c_{qi} = cluster_j \\ path_{i\ j},\ otherwise \end{array}\right\}$, where ρ, 0 ≤ ρ ≤ 1, is a user defined parameter called evaporation coefficient, Q is a positive constant and it may vary from 0 to ∞.

V. Fuzzy Enabled Clustering Scheme

5.1 Fuzzy Logic Concepts

All attribute values are converted into fuzzy values for clustering process. The fuzzy logic and fuzzy set [14] is discussed in details as follows.

Fuzzy Logic: It is based on "degrees of truth" rather than the usual "true or false" (1 or 0). It has been extended to handle the concept of partial truth. The truth values between "completely true" and "completely false". Fuzzy logic uses the whole interval between 0 and 1 to describe human reasoning.

Fuzzy Set: Fuzzy Set Theory was formalized by Professor Lofti Zadeh at the University of California in 1965. A set without a crisp boundary that is an element either "belong to a set" or "not belong to a set". The characteristic function of a fuzzy set is allowed to have values between 0 and 1.

The data-clustering scheme is designed by integrating ACO with DWC. The advantage of DWC is maintained the local consistency and global connectivity factors. The ACO with DWC is enhanced using the fuzzy logic technique to improve the accuracy. The fuzzy enhancement is done in two stages, they are: i) Distance estimation function is enhanced with fuzzy models to handle uneven data distributions and ii) fuzzy relationship analysis model is used to enhance the ant colony-clustering algorithm. The method is designed with dynamic distance measure and enhanced with fuzzy enabled ant colony clustering models, which contains four stages such as (i) Fuzzification phase (ii) Fuzzy based distance estimation (iii) Clustering process and (iv) Cluster analysis phases. These four stages as discussed in detail as follows:

(i) To obtain fuzzy set for each transaction the attribute weights are converted by fuzzification phase. Here the analysis of distance as well as global connectivity is carried out using fuzzy weight values with the range of 0 to 1. The distribution of weight value is not same for all the instance in the datasets; it may vary within the range of 0 to 1. The fuzzification process removes the overhead for calculating the distance in uncertain data distributions and handles uneven transaction distribution.

(ii) Fuzzy logic techniques are used to improve the distance estimation process. Fuzzy logic is applied to the support and distance estimation process. All the attribute weight values are converted into fuzzy weight values. The attribute weight values for each transaction are passed into the fuzzification process. The distance measure estimates the distance with global connectivity factors. All the transaction analysis is carried out with fuzzy enabled weight values.

(iii) The clustering process is done with the dynamic distance based ant colony clustering algorithm and fuzzy enabled ant colony clustering algorithm to improve the cluster accuracy. The fuzzy enabled ant colony clustering algorithm is enhanced with weight value to find the accurate distance measure as well as to improve the cluster accuracy.

(iv) Cluster accuracy is analyzed in the cluster analysis phase. In sixth section, the accuracy of proposed models are discussed in detail. The results are updated using the actual and fuzzified attribute values. Global relationship is used in the system.

The comparison process is performed with fuzzy weights. The precision/recall and fitness measures are used in the cluster analysis process. This progression has been classified into four processes such as (i) Distance Analysis (ii) Fuzzification Process (iii) Ant Colony Clustering and (iv) Fuzzy Ant Colony Clustering

5.2 Distance Analysis

Distance analysis is performed to estimate transaction relevancy. Local and global distance estimation schemes are used in the system. Local distance is estimated using the current transaction information only where as the transaction details and support information are used to estimate the global distance.

5.3. Fuzzification Process

Fuzzification comprises the process of transforming crisp values into grades of membership for linguistic terms of fuzzy sets. The membership function is used to associate a grade to each linguistic term. Fuzzy weight is used for the distance estimation and also calculating Support value. Fuzzy model is used to assign weights in a range of 0 to 1. Transaction weights are converted into fuzzy based weights. First, the data will be imported, and then it will be cleaned, then the same will be converted into fuzzified values based on three ranges.

(i) C_z - Values lies between 0 to 4 (a = 0, b = 2, c = 4) - if the data less than either 0 or 3, C_z is set to zero – if the data lies between 0 to 2, C_z will be updated using this formula ((data - a) / (b - a)). If the data lies between 2 to 4 then the C_z will be updated using C_z = ((c - data) / (c - b)).

(ii) C_0 - Values lies between 2 to 8 (a = 2, b = 5, c = 8) - if the data less than either 4 or 8, C_0 is set to zero - if the data lies between 2 to 5, C_0 will be updated using this formula C_0 = ((data - a) / (b - a)). If the data lies between 5 to 8 then the Co will be updated using C_0 = ((c - data) / (c - b)).

(iii) C_b - Values lies between 6 to 10 (a = 6, b = 8, c = 10) - if the data less than either 6 or 10, C_b is set to zero - if the data lies between 6 to 8, C_b will be updated using this formula C_b = ((data - a) / (b - a)) - if the data lies between 8 to 10 then the C_b will be updated using C_b = ((c - data) / (c - b)). Figure 2 shows the diagrammatic representation of fuzzification.

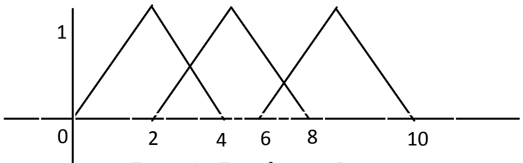

Figure 2. Fuzzification Process

5.4. Ant Colony Clustering

Ant Colony Optimization algorithm (ACO) [6] is used to optimize good and shortest pathway through pheromone trial updating. The static and dynamic optimization problem has been solved by using ACO. Ants' uses its intelligent behavior [4] to find the optimal path. Marco Dorigo in 1992 proposed that ACO [10] is a probabilistic technique for solving combinatorial problems through graphs. According to him finding the optimal path is solved by utilizing Ants' path searching behavior [11]. The optimization method is used for clustering process where the transaction weights of the optimal paths are used. Figure 3 shows ACO algorithm for clustering the breast cancer data set.

```
ACO Algorithm for Clustering Breast Cancer Data Set
Input     :  Breast Cancer Dataset, Cluster Count
Output : Clusters based on Cluster Count

Step 1    :  To find the proper food source, the ants traverse around the colony.
Step 2    :  After finding the food source it returns to nest.
Step 3    : While travelling it deposits some amount of Pheromone throughout the path.
Step 4    : With the help of pheromone deposition the follower ant reach the destiny.
Step 5    : This transaction will make strengthen the deposition of the pheromone.
Step 6    : This strengthens the route of the ant
Step 7    : In mean time the amount of pheromone will evaporate in each traversal.
Step 8    : If there are two routes to reach the same food source the ant find the shortest route between food
             and nest with the help of pheromone updating.
```

Figure 3. Ant Colony Optimization (ACO) Algorithm

5.5. Fuzzy Ant Colony Clustering

Clustering process is performed using fuzzy weights. Fuzzy weights based distance is used for the relevancy estimation. Global distance is used for the clustering process. Fuzzy relation is integrated with the ant colony clustering model is called FACO [12]. The Algorithm for FACO is presented in Figure 4.

```
FACO Algorithm for Clustering Breast Cancer Data Set
Input    : Fuzzified Breast Cancer Dataset, Cluster Count
Output : Clusters based on Cluster Count

Step 1    :  Importing and transferring dataset into an understandable format.
Step 2    : The transferred dataset will be cleaned through data cleaning process using Aggregation Function
             (Average).
Step 3    : The attribute support details will be calculated which  Consists of attribute value with
              corresponding count and support value.
Step 4    : The transaction support value will be calculated which consists of attribute name with
              corresponding attribute value and support value.
Step 5    : The support distance details for each attribute will be calculated.
Step 6    : The cleaned attribute value was converted into fuzzy values using fuzzification process.
Step 7    : The cluster count will be selected as per user needs then the details about the particular cluster are
             also make out.
Step 8    : The cleaned Fuzzified values will be clustered using Fuzzy Ant Colony clustering process.
```

Figure 4. Fuzzy Ant Colony Optimization (FACO) Algorithm

5.6 Weighted Fuzzy ACO

Weighted Fuzzy Ant Colony Optimization (WFACO) [15] can find an accurate distance and also increase the cluster accuracy. By adjusting the weights, it is able to control the growth or decay of the clusters. If the weight of a cluster increases, data points are more likely to be grouped in other clusters. Similarly, decreasing the weight helps to increase the population of a cluster. Each attribute value for all records are updated with corresponding fuzzified weight value. The Algorithm of Weighted Fuzzy Ant Colony Clustering Algorithm is projected in Figure 5. The weight value is formally defined as

$$W_j = \sum_{A_i \in J} A_i$$ where j = 1, 2, 3, …, n and i = 1, 2, 3, …, 10.

WFACO Algorithm for Clustering Breast Cancer Data Set

Input : Weighted Fuzzified Breast Cancer Dataset, Cluster Count
Output : Clusters based on Cluster Count

Step 1 : Importing and transferring dataset into an understandable format.
Step 2 : The transferred dataset will be cleaned through data cleaning process using Aggregation Function (Average).
Step 3 : The attribute support details will be calculated which consists of attribute value with corresponding count and support value.
Step 4 : The transaction support value will be calculated which consists of attribute name with corresponding attribute value and support value.
Step 5 : The support distance details for each attribute will be calculated.
Step 6 : The cleaned attribute value was converted into fuzzy values using fuzzification process.
Step 7 : Each attribute should be assigned with weight value. The weight values are calculated by adding the three different ranges of fuzzification values for each attribute. For example
WCT = Cz + Co + Cb.
Step 8 : All the fuzzy attribute values are converted into weighted fuzzy values using the step 7.
Step 9 : The cluster count was selected as per user needs then the details about the particular cluster are also making out.
Step 10 : The cleaned Fuzzified weighted values are clustered using Weighted Fuzzy Ant Colony Clustering process.

Figure 5. Weighted Fuzzy Ant Colony Optimization (WFACO) Algorithm

VI. Experimental Analysis

The algorithms described in previous section are implemented using JAVA. Brest Cancer Data set is taken from the UCI (University of California, Irwin) machine learning repository and used to conduct experiment in order to analyse the performance of the proposed approach. Totally it contains 1000 instance and 11 attribute including class attribute. This data set provides information about the breast cancer patient diagnosis information. The class information and associated symptom details are provided in the dataset. The dataset contains some noise records. Noise elimination process is performed on the data sets by using precision and recall methods.

Table 1. F-Measure values for all Cluster Count

F-measure Analysis on Clustering Techniques					
S. No.	Cluster Count	Transactions	ACO	FACO	WFACO
1	2	200	0.715	0.931	0.942
		400	0.742	0.946	0.965
		600	0.761	0.952	0.974
		800	0.784	0.977	0.989
		1000	0.797	0.998	0.994
2	3	200	0.718	0.927	0.946
		400	0.739	0.941	0.963
		600	0.758	0.956	0.979
		800	0.779	0.972	0.986
		1000	0.795	0.991	0.998
3	4	200	0.723	0.931	0.948
		400	0.736	0.943	0.967
		600	0.754	0.959	0.985
		800	0.781	0.969	0.989
		1000	0.798	0.987	0.996
4	5	200	0.727	0.935	0.951
		400	0.744	0.944	0.972
		600	0.761	0.963	0.985
		800	0.786	0.969	0.992
		1000	0.801	0.995	0.997

The parameter F-measure and entropy are used to evaluate the performance of the proposed clustering algorithms. The measurement F-measure represents the cluster accuracy information. Precision and recall values are used in the F-measure estimation process. Entropy is used to analyze the distance interval between clusters.

The inter cluster intervals are analyzed using entropy measures. The F-Measure values of various cluster count is presented in Table 1. The Entropy values of various cluster count is projected in table 2. The average F-Measure value of each cluster count for the entire dataset is depicted in table 3. The average Entropy value of each cluster count for the entire dataset is depicted in table 4. From figure 6, it concludes that the cluster count 5 works effectively and yields more accuracy compare to other cluster count in F-Measure Analysis. From Figure 7, it is observed that the cluster count 5 works effectively and yields more accuracy compare to other cluster count in Entropy Analysis.

Table 2. Entropy values for all Cluster Count

Entropy Analysis on Clustering Techniques					
S.No	Cluster Count	Transactions	ACO	FACO	WFACO
1	2	200	0.879	0.945	0.983
		400	0.882	0.959	1.017
		600	0.894	0.967	1.061
		800	0.911	0.981	1.147
		1000	0.923	0.999	1.192
2	3	200	0.875	0.942	0.981
		400	0.886	0.956	1.012
		600	0.898	0.963	1.065
		800	0.908	0.984	1.142
		1000	0.919	0.996	1.187
3	4	200	0.871	0.949	0.978
		400	0.891	0.95	1.056
		600	0.902	0.959	1.091
		800	0.912	0.987	1.178
		1000	0.923	1.011	1.192
4	5	200	0.871	0.939	0.985
		400	0.892	0.96	1.021
		600	0.903	0.968	1.056
		800	0.916	0.985	1.124
		1000	0.925	0.999	1.178

Table 3. Average F-Measure value of each cluster count for the entire dataset

Average Entropy Analysis on Clustering Techniques				
S.No	Cluster Count	ACO	FACO	WFACO
1	2	0.8978	0.9702	1.0800
2	3	0.8972	0.9682	1.0774
3	4	0.8998	0.9702	1.0728
4	5	0.9014	0.9712	1.0990

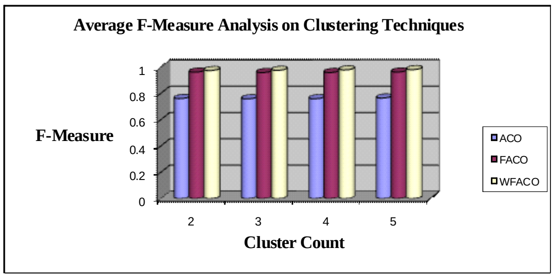

Figure 6. Average F-Measure Analysis of each cluster count for the entire dataset

Table 4. Average Entropy value of each cluster count for the entire dataset

Average F-measure Analysis on Clustering Techniques				
S.No	Cluster Count	ACO	FACO	WFACO
1	2	0.7598	0.9608	0.9728
2	3	0.7578	0.9574	0.9744
3	4	0.7584	0.9578	0.9778
4	5	0.7638	0.9612	0.9794

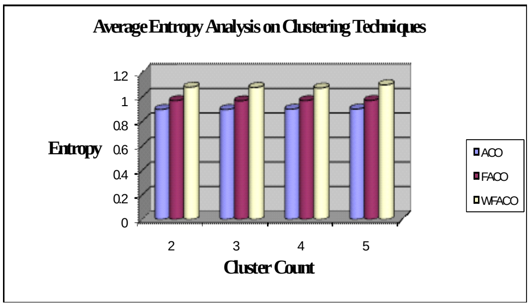

Figure 7. Average Entropy Analysis of each cluster count for the entire dataset

VII. CONCLUSION

The Proposed approach uses data connectivity and an improved formula for calculating the distance named DWC instead of Euclidean Distance. DWC reflects not only the local consistency but also the global connectivity between objects. It also overcomes the shortcoming of Euclidean Distance in data clustering. Then, the ant colony-clustering algorithm is improved by using DWC and fuzzy logic concepts. Then the Fuzzy ACO has extended with weight value to find the accurate distance measure. The experimental results for Breast cancer data set show that the improved algorithm can discover clusters with arbitrary shape and is better than the clustering effect of earlier techniques. The limitation of this work is not reach to 100% accuracy. Further this work can be extended by incorporating rough set model to increase the accuracy clustering model for the breast cancer dataset.

REFERENCES

[1] Jiawei Han and Micheline Kamber.Data Mining Concepts and Techniques, San Francisco: Morgan kaufmann, 2006, pp.383.

[2] Deneubourg JL, Goss S, et a1. “The dynamics of collective sorting: robot-like ant and ant-like robot,”. In: M eyer JA, Wilson SW ed. Proceedings first conference on simulation of adaptive behavior: from animals to animats. Cambridge, MA: MIT Press, 1991, pp.356–365.

[3] Shiyong Li, Baojiang Zhao. “Ant Colony Clustering Algorithm,” Measurement & Control, vol. 11, NO.15, 2007, pp.159–1592, 2007.15(11):1590–1592.

[4] Parag M. Kanade and Lawrence. Hall, “Fuzzy Ants and Clustering, IEEE Transactions on Systems,” man, and Cybernetics-part a: systems and humans, vol. 37, no. 5, September 2007, pp. 758–769.

[5] Xinbin Yang, Jinggao Sun and Dao Huang, “A New Clustering Method Based on Ant Colony Algorithm,” Proceeding of the 4th World Congress on Intelligent Control and Automation June 10–14, 2002, pp.2222-2226.

[6] Jian Gao. “Cluster Analysis Based on Parallel Ant Colony Adaptive Algorithm,” Computer Engineering and Application, vol. 25, 2003, pp.78–79, 2003.25:78–79.

[7] Yan YANG and Fan Jin, “Mohamed Kamel.Clustering Combination Based on Ant Colony Algorithm,” Journal of the China Railway Socity, vol. 4, No. 26, 2004, pp.6–69.

[8] Maoguo Gong and Liefeng Bo. “Density-Sensitive Evolutionary Clustering,” The 11th Pacific- Asia Conference on Knowledge Discovery and Data Mining, Springer-Verlag Berlin Heidelberg ,2007, pp.507–514.

[9] Shanfei Li, Kewei Yang, Wei Huang, Yuejin Tan, “An Improved Ant-Colony Clustering Algorithm Based On The Innovational Distance Calculation Formula” Third International Conference on Knowledge Discovery and Data Mining, ISBN: 978-0-7695-3923-2, 2010

[10] Julia Handl and Joshua Knowles. “An Evolutionary Approach to Multiobjective Clustering,” IEEE Transactions on Evolutionary Computation, vol. 11, no. 1, Feb.2007, pp.60.

[11] Miguel A.Sanz-Bobi and Mario Castro “IDSAI: A Distributed System for Intrusion Detection Based on Intelligent Agent” 5th International Conference on Internet Monitoring and Protection, IEEE, 2010.

[12] Amin Einipour “A Fuzzy-ACO Method for Detect Breast Cancer” Global Journal of Health Science October 2011.

[13] Niloofar Maleki, MohammadReza DehghaniMahmoudAb, Mehdi Bahrami, "Optimization of Ant Colony Routing for Ad Hoc Communication Networks", International Journal of Soft Computing and Software Engineering [JSCSE], Vol. 2, No. 4, pp. 36-45, 2012, Doi: 10.7321/jscse.v2.n4.4

[14] Mehdi Bahrami, Mohammad Bahrami, An overview to Software Architecture in Intrusion Detection System, International Journal of Soft Computing And Software Engineering (JSCSE), ISSN: 2251-7545, Vol.1,No.1, 2011.

[15] S.Nithya and R.Manavalan, “An Ant Colony Clustering Algorithm Using Fuzzy Logic”, International Journal of Soft Computing and Software Engineering (JSCSE), e-ISSN: 2251-7545, DOI: 10.7321/jscse, Vol: 2, No: 5. 2012.

Ms. NITHYA .S completed her B.Sc Information Technology in C.S.I college of Arts and Science and M.Sc Computer Science and Information Technology in Yadava College of Arts and Science with Distinction. She pursues M.Phil in Computer Science in K S Rangasamy College of Arts and Science. Her areas of interest are Data Mining and Soft Computing.

Mr MANAVALAN R Obtained M.Sc., Computer Science from St.Joseph’s College of Bharathidasan University,Trichy, Tamilnadu, India, in the year 1999, and M.Phil., in Computer Science from Manonmaniam Sundaranar University, Thirunelveli, Tamilnadu, India in the year 2002. He pursues Ph. D. in Computer Science, at Periyar University, Salem. He works as Asst. Prof & Head, Department of Computer Science and Applications, KSR College of Arts and Science, Thiruchengode, Namakwa, Tamilnadu, India. His areas of interest are Medical Image Processing and Analysis, Soft Computing, Pattern Recognition and Theory of Computation.