ISSN:0975-9646

Performance Analysis of Clustering Based on Fuzzy System

K.Vijayakumari[#1], M. Preetha [*2], K.Velusamy[#3]

[#]*PGP College of Arts and Science - DIM3 Technology and Solutions*
Namakkal, Comibatore, India

[*]*PGP College of Arts and Science*
Namakkal, India

***Abstract*— the Bee Colony Optimization algorithm is a population based swarm intelligent family, it has to solved local optima problem and overwrite the demerit of Particle swarm optimization algorithms. BCO are inspired by the principles of natural biological behaviour and circulated communal behaviour of social colonies has shown superiority in production with complex optimization problems. In this paper, Fuzzy Bee Colony Optimization (FBCO) was wished-for that algorithm furnished incredibly shining effect match up to with other active algorithms such as Fuzzy C- Means and Fuzzy Particle Swarm Optimization.**

***Keywords*— Clustering, Optimization, Fuzzy C-Means, Fuzzy Particle Swarm Optimization, Fuzzy Bee Colony Optimization,**

I. INTRODUCTION

The user can extract the unfamiliar pattern from the bulky set of data for trade as well as real time applications by using fundamental concept of data mining [3]. The data mining is valuable tools for data analysis and it is a computational intelligence discipline is very useful and new knowledge discovery and autonomous decision making.

Data mining is one of the most vital investigate field that are due to the development of both computer hardware and software technologies, Clustering is essential in data analysis and data mining appliance.

Fig 1.1 Data Mining Process

Classification and prediction are two shape of data analysis that can be used to take out models. Such examination can helps to make available better considerate of the data at large. Classification predicts categorical (discrete, unordered) labels, prediction models continuous valued functions. Many Classification and prediction methods have been proposed by researchers in machine learning, pattern recognition, and statistics.

A. Applications

- Financial Data Analysis
- Retail Industry
- Telecommunication Industry
- Biological Data Analysis
- Other Scientific Applications
- Intrusion Detection

B. Requirements of Clustering in Data Mining

- Scalability
- Dealing with dissimilar types of attributes
- invention of clusters with arbitrary shape
- Minimal requirements for domain knowledge to determine input parameters
- Able to deal with noise and outliers
- Insensitive to order of input records
- High dimensionality
- Interpretability and usability

II. UNSUPERVISED CLASSIFICATION

In 1968, Jar dine and Sibson were developed clustering method that is an unsupervised technique. The aim is to find out groups of analogous instances within the dataset exclusive of category label. The aspiration of clustering [17] is to collection of sets of bits and pieces into classes such that analogous objects are positioned in the same cluster while dissimilar objects are in break up into various clusters [8].

It helps the addict to appreciate the expected grouping or construction in a data set. Clustering is the unsupervised classification without predefined classes [2]. A high-quality clustering method will produce high quality clusters in which the intra-class similarity is high and the inter-class similarity is low [7]. Clustering seek at representing large datasets by a fewer number of prototypes or clusters. It brings simplicity in modeling data and thus plays a central role in the procedure of knowledge innovation and data mining.

In clustering, there are so many a variety of numbers of algorithms such as hierarchical, partitioning, grid and density based algorithms.

Lotfi Zadeh was developed fuzzy theory and integrated with it into clustering technique which is called fuzzy clustering. Fuzzy algorithms can assign data entity incompletely to multiple clusters. The degree of membership in the fuzzy clusters depends on the familiarity of the data object to the cluster centers. The majority of well-liked fuzzy clustering algorithm is Fuzzy C-Means (FCM) which developed by Bezdek in 1974 and meanwhile it is widely used [6], [12].

Fuzzy C-Means clustering [4] is a capable algorithm, FCM occupation as random selection in center points construct iterative evolution falling into the local optimal solution without difficulty. So many investigator used for solving this optimization problem of evolutionary algorithms such as Genetic Algorithm (GA), Simulated Annealing (SA), Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO) have been successfully applied. The following sections are described about Fuzzy C-Means [9], Fuzzy Particle Swarm Optimization [10] and Fuzzy Bee Colony Optimization.

A. *Fuzzy C-Means Algorithm (FCM)*

The Fuzzy C-Means algorithm [1] is an iterative algorithm that stumbles on clusters in order and which uses the concept of fuzzy membership instead of passing on a pixel to a single cluster, each pixel will have extraordinary membership values on each cluster.

FCM is panel of place of n substance $O = \{O_1, O_2, ...O_n\}$ in Rd dimensional hole into c (1 < c < n) fuzzy clusters with $z = \{z_1, z_2, ..., z_n\}$ cluster centers or centroids. The fuzzy clustering of objects is explained by a fuzzy matrix μ with n rows and c columns in which n is the quantity of data substance and c is the figure of clusters. μ_{ij}, the component in the i_{th} row and j_{th} column in μ, point out the degree of association or membership function of the i_{th} object with the j_{th} come together. The typescript of μ is as go behind:

$$\forall_i \ [0, 1] \quad \forall_i = 1, 2 \ldots n \quad \forall_j = 1, 2 \ldots c \quad (2.1)$$

$$\sum_{j=1}^{c} \mu_{ij} = 1 \quad \forall_i = 1, 2 \ldots n \quad (2.2)$$

$$o < \sum_{j=1}^{c} \mu_{ij} < n \quad \forall_j = 1, 2 \ldots c \quad (2.3)$$

In FCM has that objective function is minimize the Eq. (2.4):

$$J_m = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij} d_{ij} \quad (2.4)$$

Where

$$d_{ij} = | \ o_i - z_j \ | \quad (2.5)$$

In which, m (m>1) is a scalar phrase the weighting exponent and controls the fuzziness of the resulting clusters and d_{ij} is the Euclidian distance from object $z = \{z_1, z_2, ..., z_n\}$ to the cluster center z_j. The z_j, centroids of the j_{th} cluster, is get clutch of using under equation

$$z_j \frac{\sum_{i=1}^{n} \mu_{ij}^{m} o_i}{\sum_{i=1}^{n} \mu_{ij}^{m}} \quad (2.6)$$

Algorithm 1: Fuzzy C Means

Step 1: Initialize the membership function values μ_{ij}

Step 2: Compute the cluster centers z_j

Step 3: Compute Euclidian distance d_{ij}

Step 4: Update the membership function μ_{ij},

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}}{d_{ik}})^{\frac{2}{m-1}}} \quad (2.7)$$

Step 5: If not converged, go to step 2.

B. *Particle Swarm Optimization (PSO)*

Optimization procedure characterizes a global function and tries to optimize its value by traversing the search space. Particle swarm optimization (PSO) is a [16] population based stochastic optimization method motivated by bird flocking and fish schooling in the beginning of designed and introduced by Kennedy and Eberhart in 1995 and is based on iterations/generations [5]. The algorithmic of PSO establish with a population of particles whose positions represent the potential solutions for the studied problem, and velocities are arbitrarily initialized in the search space. In each iteration the search for optimal position is personal best position and global best position. The personal best position, pbest, is the best location the particle has vacation and gbest is the best place the swarm has visited since the first time step [11]. A particle's velocity and position are updated as follows.

$$V(t+1) = w. V(t). c_1 r_1 (pbest(t) - X(t) + c_2 r_2 (gbest(t) - X(t)) \quad k = 1,2 \ldots P \quad (2.8)$$

$$X(t+1) = X(t) + V(t+1) \quad (2.9)$$

where X and V are position and velocity of particle respectively, w is inertia weight, c1 and c2 are positive constants, called acceleration coefficients which control the influence of pbest and gbest on the search process, P is the number of particles in the swarm, r1 and r2 are random values range between 0 and 1.

A. Fuzzy particle swarm optimization for fuzzy clustering

A personalized particle swarm optimization for TSP called fuzzy particle swarm optimization (FPSO) estimated by Peng et al. In which routine the position and velocity of particles redefined to symbolize the fuzzy family member between variables. In this fragment we describe this method for fuzzy clustering problem [10].

In FPSO algorithm X , the position of particle, shows the fuzzy relation from set of data objects, $o = \{o_1, o_2, ..., o_n\}$ to set of cluster centers, $z = \{z_1, z_2, ..., z_n\}$, X Can be expressed as follows:

$$X = \begin{pmatrix} \mu_{11} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nc} \end{pmatrix} \quad (2.10)$$

In which μ_{ij} is the membership function of the i_{th} object with the j_{th} cluster with constraints stated Eq. (2.1) and Eq. (2.2) therefore we can see that the position matrix of each particle is the same as fuzzy matrix μ in FCM algorithm. Also the velocity of each particle is stated using a matrix with the size *n* rows and *c* columns the elements of which are in range between -1 and 1. We get the Eq. (2.11) and Eq. (2.12) for updating the positions and velocities of the particles based on the matrix.

$$V(t+1) = w \otimes V(t) \oplus c_1 r_1 \otimes (pbest(t) \ominus X(t) \oplus c_2 r_2 \otimes (gbest(t) - \ominus X(t)) \quad k=1,2\ldots P \tag{2.11}$$

$$X(t+1) = X(t) \oplus V(t+1) \tag{2.12}$$

After updating the position matrix, it may violate the constraints given Eq. (2.1) and Eq. (2.2). So it is necessary to normalize the position matrix. First we make all the negative elements in matrix to become zero. If all elements in a row of the matrix are zero, they need to be re-evaluated using series of random numbers within the interval between 0 and 1, and then the matrix undergoes the following transformation without violating the constraints:

$$Xnormal = \begin{pmatrix} \mu_{11} / \sum_{j=1}^{c} \mu_{1j} & \cdots & \mu_{1c} / \sum_{j=1}^{c} \mu_{1j} \\ \vdots & \ddots & \vdots \\ \mu_{n1} / \sum_{j=1}^{c} \mu_{nj} & \cdots & \mu_{nc} / \sum_{j=1}^{c} \mu_{nj} \end{pmatrix} \tag{2.13}$$

In FPSO algorithm the same as other evolutionary algorithms, we need a function for evaluating the generalized solutions called fitness function. In this paper below equation is used for evaluating the solutions.

$$f(X) = \frac{K}{J_m} \tag{2.14}$$

Therein K is a constant and J_m is the objective function of FCM algorithm (Eq. (2.4)). The smaller is J_m, the better is the clustering effect and the higher is the individual fitness *f (X)*. The FPSO algorithm for fuzzy clustering problem can be stated as follows:

Algorithm 2: Fuzzy PSO for fuzzy clustering

Step 1: Initialize the parameters including population size P, c1, c2, w, and the maximum iterative count.
Step 2: Create a swarm with P particles (X, pbest, gbest and V are n× c matrices).
Step 3: Initialize X, V, pbest for each particle and gbest for the swarm.
Step 4: Calculate the cluster centers for each particle
Step 5: Calculate the fitness value of each particle
Step 6: Calculate pbest for each particle.
Step 7: Calculate gbest for the swarm.
Step 8: Update the velocity matrix for each particle
Step 9: Update the position matrix for each particle
Step 10: If terminating condition is not met, go to step 4.

The termination condition in proposed method is the maximum number of iterations or no improvement in gbest in a number of iterations.

C. *Bee Colony Optimization (BCO)*

The bee colony optimization [15] is a meta heuristic be in the right place to the identify of temperament inspired population based algorithms and that are inspired by a variety of biological and natural process that natural classification have become vital sources of thoughts and reproduction for development of a choice of artificial system [13], [14].

A. *Fuzzy Bee Colony Optimization*

The bee colony optimizations incorporate with fuzzy set theory is called Fuzzy Bee Colony Optimization (FBCO) which all constraint are based Fuzzy PSO such as 2.13 and fitness value is based on 2.14.

Step 1 Initialization Phase

- Initialize the cluster number c, the real number m, the size of the population of N,
- Generate initial population *zi,*
- Calculate the membership matrix by randomly
- Evaluate the population
- Set cycle to 1

Step 2 Repeat

Step 3 for Employed Bee

- Produce new solution *ui*
- Calculate the membership matrix
- Calculate the fitness
- Apply the greedy selection process
- Calculate the probability values *pi* for the solutions

Step 4 for Onlooker Bee and Scout Bee

- Choose a solution *zi* depending on *pi*
- Produce new solution *ui*
- Calculate the membership matrix
- Calculate center and distance value
- Calculate the fitness using
- Apply the greedy selection process
- If there is abandoned solution then replace that solution with a new randomly produced
- solution for the scout
- Assign cycle to cycle + 1
- Memorize the best solution (best cluster centers) achieved yet
- Calculate objective value

Step 5 Until cycle reach converges

The Bee colony optimizations exertions in a self organized and decentralized way and therefore correspond to a high-quality basis for parallelization. It also poses a capability to go on away from becoming attentive in local minima.

B. *Comparative between FPSO and FBCO*

TABLE 2.1 PERFORMANCES ANALYSIS OF OBJECTIVE VALUES

S.No	FPSO	FBCO
1	Weakness of search with local search	Strength is local and global search
2	It has a slow convergence rate,	Fast convergence

III. RESULT AND EXPERIMENTAL ANALYSIS WITH DISCUSSIONS

The various research papers have been used clustering method by using Bee colony optimizations for solving many problems. In this segment in attendance the experimental evidences and results that were made on a number of standard datasets, and the comparisons that were finished with other applicable works.

The UCI database, which is a distinguished catalogue repository are used to estimate the performance of the wished-for algorithm. Fisher's Iris plants database number of object is 150, dimension is 4 and which cluster is K = 3. This dataset are implemented by using MATLAB.

TABLE 3.1 PERFORMANCES ANALYSIS OF OBJECTIVE VALUES

Methods	Best	Average	Worst
FCM	64.63	62.49	61.12
FPSO	62.33	61.16	60.76
FBCO	**59.33**	**57.29**	**56.93**

The above table 3.1 give a picture of make out the performance of every used language for fuzzy clustering problem and from that table be familiar with the FBCO is best recital based on their objective values.

Fig 3.1 Depict to identify performance analysis of all algorithms and every algorithms have a lot demerit and merits, FBCO overwrite demerit of existing and produced better performance than others.

IV. CONCLUSIONS

In this paper, an optimization algorithm inspired by the natural foraging behaviour of honey bees called Bee Colony Optimization is built-in with fuzzy theory. Among them, the Fuzzy Bee Colony Optimization is making available to well-organized conclusion for fuzzy clustering in data mining match up to with other on hand algorithms. This natural procedure of work provides a number of ways for solving the real world problems more powerfully and rapidly with accuracy.

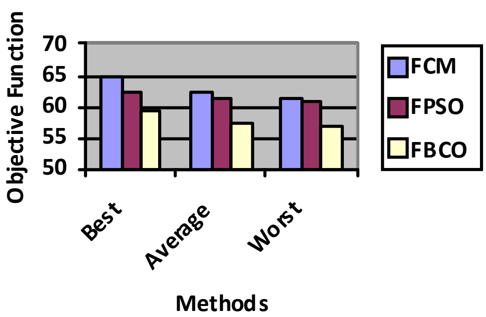

Fig 3.1 Performance analysis of objective values

REFERENCES

[1] CHIH-TANG CHANG, JIM Z. C. LAI AND MU-DER JENG, "A Fuzzy K-means Clustering Algorithm Using Cluster Center Displacement", JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 27, 995-1009 (2011).

[2] Pritesh Vora, Bhavesh Oza, "A Survey on K-mean Clustering and Particle Swarm Optimization", International Journal of Science and Modern Engineering (IJISME) ISSN: 2319-6386, Volume-1, Issue-3, February 2013.

[3] Tiwari, Manu Bhai Jha, and OmPrakash Yadav, "Performance analysis of Data Mining algorithms in Weka", IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661, ISBN: 2278-8727 Volume 6, Issue 3 (Sep-Oct. 2012), PP 32-41.

[4] Amreen Khan, Prof. Dr. N.G.Bawane and Prof. Sonali Bodkhe, "An Analysis of Particle Swarm Optimization with Data Clustering-Technique for Optimization in Data Mining", Amreen Khan et. al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 04, 2010, 1363-1366.

[5] Asha Gowda Karegowda and Seema Kumari, "Particle Swarm Optimization Algorithm Based k-means and Fuzzy c-means clustering", Volume 3, Issue 7, July 2013 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering.

[6] Soumi Ghosh and Sanjay Kumar Dubey, "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms",((IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No.4, 2013.

[7] Amandeep Kaur Mann (M.TECH C.S.E) and Navneet Kaur (Assistant.Professor in C.S.E) , "Survey Paper on Clustering Techniques", ISSN: 2278 – 7798 International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 4, April 2013 803.

[8] Pawan Kumar, Mr. Pankaj Verma, Rakesh Shrma, "Comparative Analysis of Fuzzy C Mean and Hard C Mean Algorithm", International Journal of information Technology and Knowledge Management July-December 2010, Volume 2.

[9] Mrs. Bharati R.Jipkate and Dr.Mrs.V.V.Gohokar, "A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms" International Journal of Computational Engineering Research, ISSN: 2250–3005IJCER, May-June 2012 Vol. 2 Issue No.3 737-739.

[10] Esmaeil Mehdizadeh, "A fuzzy clustering PSO algorithm for supplier base management", ISSN 1750-9653, England, UK International Journal of Management Science and Engineering Management Vol. 4 (2009) No. 4, pp. 311-320.

[11] Chia-Feng Juang Che-Meng Hsiao and Chia-Hung Hsu, "Hierarchical Cluster-Based Multispecies Particle-Swarm Optimization for Fuzzy-System Optimization" IEEE Transactions on Fuzzy Systems, Vol. 18, No. 1, February 2010.

[12] Chen Yanyun , Qiu Jianlin, Gu Xiang,Chen Jianping, Ji Dan, Chen Li, "Advances In Research of Fuzzy C-means Clustering Algorithm", 2011 International Conference on Network Computing and Information Security. 978-0-7695-4355-0/11 $26.00 © 2011 IEEE DOI 10.1109/NCIS.2011.104.

[13] Punam Priti Pradhan, Debahuti Mishra, Sashikala Mishra and Kailash Shaw, "Artificial Bee based Optimized Fuzzy c-Means Clustering of Gene Expression Data", International Journal of Computer Trends and Technology (IJCTT) – Volume 4 Issue 5–May 2013.

[14] M.Nandhini, K.Mohana Prasad, "An Effective Method for Clustering Using an Algorithm Stimulated By the Decision Making Processes of Bee Swarms", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-3, Issue-3, February 2014.

[15] Baris Yuce , Michael S. Packianather , Ernesto Mastrocinque , Duc Truong Pham and Alfredo Lambiase, "Honey Bees Inspired Optimization Method: The Bees Algorithm", Insects 2013, 4, 646-662; doi:10.3390/insects4040646, ISSN 2075-4450.

[16] K. Velusamy and R.Manavalan , "Performance Analysis of Unsupervised Classification based on Optimization," International Journal of Computer Applications (0975 – 8887, Volume 42– No.19, March 2012.

[17] K. Velusamy, A.Sakthivel, and M. Jayakeerithi "Studies on clusterin and fuzzy clustering", International Journal of Computer Science and Information Technologies vol . 5(4), 2014, 5231-5232, ISSN: 9775-9646.