PERFORMANCE ANALYSIS OF OPTIMIZATION TECHNIQUES BY USING CLUSTERING

K. Vijayakumari[1], Dr V. Baby Deepa[2]

[1]Research Scholar, Department of Computer Science, Government Arts College, Karur, Tamilnadu, India
[2]Assistant Professor, Department of Computer Science, Government Arts College, Karur, Tamilnadu, India

---

Abstract - *the Bee Colony Optimization algorithm shows the new path in the field of swarm intelligent family, it has the capability to avoid the local minima problem and it overcomes the difficulties in Particle swarm optimization algorithms. BCO are motivated by the principles of natural genetic activities and spread common behaviour of social colonies has shown authority in production with complex optimization problems. In this paper, Fuzzy Bee Colony Optimization (FBCO) algorithm gives better result to match up with other swarm algorithms such as Fuzzy C- Means and Fuzzy Particle Swarm Optimization.*

***Key Words*: Clustering, Optimization, Fuzzy C-Means, Fuzzy Particle Swarm Optimization, Fuzzy Bee Colony Optimization**

1. INTRODUCTION

The data will available in unusual format for that suitable action to be taken by the user, by preserve the data and by making good decision. Consumers will essential to retrieve the data from database by making better decision, this refer to Data mining or Knowledge Discovery process [1]. The data mining [3] is important tools for data analysis and it is a computational intelligence discipline is very practical and new knowledge discovery and independent decision making.

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

In Data Mining there are two forms of data analysis there are classification and prediction. This assessment can help to make obtainable enhanced considerate of the data at huge. Classification predicts definite (discrete, unordered) labels, prediction models permanent valued functions.

A. Applications of clustering

- Biology, computational biology and bio informatics
- Medicine
- Business and marketing
- World wide web
- Pattern Recognition
- Image segmentation
- Spatial Data analysis

The aim of Clustering is to represent large datasets by a smaller number of clusters. It gets simplicity in demonstrating information and thus acts as essential role in the method of knowledge discovery and data mining. The result of the clustering process and competence of its field request are generally determined through algorithms [2], there are various algorithms which are used to solve this problem.

2. UNSUPERVISED CLASSIFICATION

Unsupervised clustering technique were developed by Jar dine and Sibson in 1968. Its goal to discover group of similar instance within the dataset limited of group label. The aspiration of clustering is to collection of sets of bits and pieces into classes such that similar objects are located in the same cluster while dissimilar objects are in break up into various clusters [8].

Clustering techniques are generally unsupervised classification without predefined classes it can be used to classify data into groups based on resemblance along with the being data objects. An expert clustering method will create high superiority clusters in which the intra-class resemblance is high and the inter-class resemblance is low [7].

One of the capable algorithms in clustering [4] is Fuzzy C-Means, its job as arbitrary collection of centre points create iterative development declining into the limited best clarification without complication.

2.1 *Fuzzy C-Means Algorithm (FCM)*

FCM is section of position 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 clustering of fuzzy objects is explained by a fuzzy matrix μ with n rows and c columns in which n is the amount of information and c is the figure of clusters. μ_{ij}, the part within 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:

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

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

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

The objective function of FCM is to reduce the Eq. (2.4):

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

$$\text{Where } d_{ij} = | o_i - z_j | \qquad (2.5)$$

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 $z = \{z_1, z_2, \ldots, z_n\}$ to the cluster center z_j. The z_j, centroids of the j_{th} cluster, is get clutch of using under equation

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

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},

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

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

2.2 Particle Swarm Optimization (PSO)

It is projected by Eberhart and kennedy in 1995. It is an optimization algorithm stimulated by the common activities of birds and used to resolve all kinds of optimization problem. Particle swarm optimization is a type of swarm intelligence algorithm where the agents known as particles attempt to optimize a fitness function by moving through a explore hole. In PSO, the inhabitant's dynamics resembles the progress of a bird's flock pointed for food, here information distribution takes place and individuals can gain from the discoveries and previous experience from all other particles. The speed of particles determines the direction and distance of particle movement, while the velocity of the particle is adjusted dynamically with the movement of its own and other particles so as to recognize optimization of the individual in the solution space of optimization. Particles renew their velocity and position through two kinds of 'best'value. One is its personal best (*p best)*, which is the location of its highest fitness value. Another is the global best (*g best*), which is the location of overall best value, obtained by any particles population. The positions and velocities of the elements are given in the equations

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

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

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

A. Fuzzy particle swarm optimization for fuzzy *clustering*

In FPSO algorithm X [11], the location of element, finds the fuzzy family member from set of data objects, $o = \{o_1, o_2, \ldots, o_n\}$ to set of cluster centers [10], $z = \{z_1, z_2, \ldots, z_n\}$, X Can be expressed as follows:

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

Where μ_{ij} is the membership task of the i_{th} point with the j_{th} cluster with condition in Eq. (2.1) and Eq. (2.2) therefore we can able to see that the location 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.

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

$$X(t+1) = X(t) \oplus V(t+1) \quad (2.12)$$

After updating the position matrix, it may go against to the conditions given in Eq. (2.1) and Eq. (2.2). So it is essential to stabilize the position matrix. First we mark zero to all the negative elements in matrix. 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 performs the following transformation with its constraints:

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

In FPSO algorithm the same as other evolutionary algorithms, we need a function for evaluating the generalized solutions called fitness function.

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

This above equation is used for solution evaluation. Here *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 every 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 every particle

Step 9: Update the position matrix for every particle

Step 10: If terminating condition not satisfied, go to step 4.

The condition in proposed method is the highest number of iterations or no development in gbest in a number of iterations.

2.3 Bee Colony Optimization (BCO)

The Bees Algorithm is a new search algorithm, first residential in 2005 by Pham DT etc.and Karaboga.D [9]. The algorithms replicate the food foraging behaviour of swarms of honey bees. In fundamental version, the algorithm performs a kind of neighbourhood search combined with random search and can be used for optimization problems. Bee Colony Optimization algorithm proposed by Teodorovic and Dell'Orco [13, 14].

Fuzzy Bee Colony Optimization

The bee colony integrate with fuzzy system 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.

Algorithm 3: Fuzzy BCO for fuzzy clustering

Step 1 Initialization Phase

- Initialize the number of cluster as c, the real number m, the population size 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 results

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 the solution is discarded then replace that solution with a new random solution for the scout
- Allocate cycle to cycle + 1
- Remember the best solution (best cluster centers) achieved yet
- Calculate objective value

Step 5 Until cycle reach converges

Table -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

3. RESULT AND EXPERIMENTAL ANALYSIS WITH DISCUSSIONS

There are many research papers have been used in Bee colony optimization with clustering methods for solving many problems. The UCI database, which is a well-known directory storehouse are used to estimate the concert of the algorithm. The number of object is 150, dimension is 4 and cluster is K = 3 in Fisher's Iris plants dataset. This dataset are implemented by using MATLAB.

Table -2: Performances analysis of statics parameter

Methods	Accuracy	Sensitivity	Specificity
FCM	0.42	0.78	0.72
FPSO	0.53	0.86	0.77
FBCO	**0.55**	**0.89**	**0.81**

The table 2 give a picture of make out the presentation of fuzzy clustering problem and from that table be well-known with the FBCO is best concert based on their objective values.

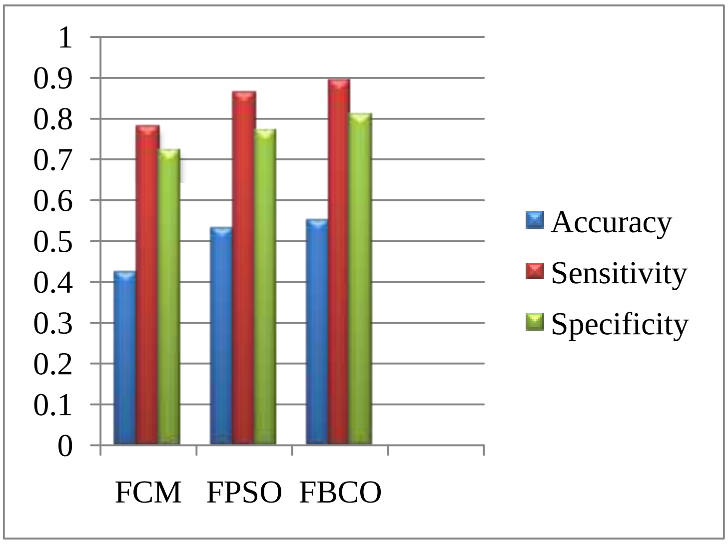

Fig 1 Performance Analysis of their Parameter

Fig 1 illustrate to categorize the performance analysis of all algorithms and each algorithm have a lot of demerits and merits, FBCO overwrite the demerit of existing and created improved performance than others.

4. CONCLUSION

In this paper, the bee colony optimization is integral with fuzzy theory, along with the Fuzzy Bee Colony Optimization is construction offered to efficient conclusion for fuzzy with clustering in data mining equivalent up with other algorithms. This normal method of effort provides a number of ways for solving the real world problems more effectively and quickly with accuracy.

REFERENCES

[1] Neelamadhab Padhy, Dr. Pragnyaban Mishra , and Rasmita Panigrahi, "The Survey of Data Mining Application And Feature Scope", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.3, June 2012

[2] Malwinder singh and Meenakshi bansal, "Survey On Clustering And Optimization Techniques To Develop Hybrid Clustering Technique", International Journal of Computer Engineering and Applications, Volume VII, Issue I, July 14 www.ijcea.com ISSN 23213469.

[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] Makhalova Elena, "Fuzzy C - Means Clustering In Matlab", The 7th International Days of Statistics and Economics, Prague, September 19-21, 2013.

[8] Ren Jingbiao,Yin Shaohong, "Research and Improvement of Clustering Algorithm in Data Mining", *2010 2nd International Conference on Signal Processing Systems (ICSPS)* 978-1-4244-6893-5/$26.00 C 2010 IEEE.

[9] Li-Pei Wong, ii Malcolm Yoke Hean Low, iii Chin Soon Chong, "Bee Colony Optimization with Local Search for Traveling Salesman Problem",www.informatik.uniheidelberg.de/groups/comopt/software/TSPLIB95.

[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] Dusan TEODOROVIC, Mauro DELL ORCO" Bee Colony Optimization – A Cooperative Learning Approach To Complex Transportation Problems" Advanced OR and AI Methods in Transportation

[14] 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.