ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 2, Issue 9, September 2013

Metaheuristics Bio inspired based optimization for data clustering problem

K.Papitha

***Abstract*—** The primary goal of data mining is to extract the knowledge from available data. It is also a form of knowledge discovery essential for solving problem in a specific domain. One of the efficient mining techniques called Clustering Analysis includes a number of different algorithm and method for grouping objects by their similar characteristics into categories. This paper explores four different bio-inspired metaheuristics in the clustering problem namely Genetic Algorithms (GA), Ant Colony Optimization (ACO), Particle Swarm optimization (PSO) and Firefly Algorithm (FA). Data mining approaches are applied in the field of medical diagnosis recently. The major class of problem in medical science involves diagnosis of disease based upon various tests. The computerized diagnostic tools are helpful to predict the diagnosis accurately. Breast cancer is one the most dangerous cancer type in the world. This disease is characterized by uncontrolled cell growth in tissue of the breast and spread through such continuous abnormal cell growth. Early detection can save a life and increase survivability of the patients. Here we propose Breast Cancer data mining using the already mentioned metaheuristics in order to improve their performance in the data clustering problem. This work proposes an algorithm for data mining called GAPF techniques (GA, ACO, PSO and FA) that focus on Breast Cancer Data Clustering that performs efficient cluster discovery with evolutionary algorithm employed to optimize the obtained result. The data set was derived from UCI machine learning repository. This proposed approach has potential applications in hospital decision-making and research such as predictive medicine.

***Index Terms*—** Breast cancer data clustering , GAPF, GA-ACO-PSO & FA technique and hospital breast cancer diagnosis.

I. INTRODUCTION

Introduction

Clustering psychoanalysis is an investigative information analysis tool which aim at sorting dissimilar data objects into group in a way with the intention of the amount of association between two matters is maximal if they be in the right place to the similar group and negligible or else. Then, clustering analysis can be used to determine structure in information with no any prior information. The resultant taxonomies have to get together the subsequent properties: homogeneity inside the clusters and heterogeneity between clusters [1][2]. Thus, it is attractive to obtain the maximum resemblance between data points into a gather and the maximum dissimilarity among data points from dissimilar clusters. One of the approach used to solve clustering evils is the use of cluster centers that are make-believe points in the investigate space. Each point is classified using Euclidean distance metric to the nearest center.Optimization is a commonly encountered mathematical problem in all manufacturing discipline. It factually means judgment the best probable/desirable explanation. Optimization problems are extensive range and many, hence methods for solve problems ought to be, a vigorous investigate topic. Optimization algorithms are able to be either deterministic or stochastic in scenery. Previous methods to solve optimization problems necessitate enormous computational hard work, which is inclined to fail as the difficulty size increase. This is the motivation for employing bio inspired stochastic optimization algorithms as computationally efficient alternatives to deterministic approach. Meta-heuristics are based on the iterative improvement of either a population of solutions (as in Evolutionary algorithms, Swarm based algorithms) or a single solution and mostly employ randomization and local search to solve a given optimization problem. Heuristic approach seems to be greater in solve inflexible and composite optimization problems, predominantly where the conventional methods fail. BIAs are such heuristics that mimics/imitates the approach of nature since many organic processes can be consideration of as process of controlled optimization. They create use of many random decisions which classify them as a particular class of randomized algorithms. Formulating a design for bioinspired algorithms involves choosing a proper symbol of difficulty; evaluate the excellence of answer using a fitness function and important operator so as to create a new set of solution. Metaheuristics, such as Genetic Algorithms (GA), Ant Colony Optimization (ACO), Particle Swarm optimization (PSO) and Firefly algorithm (FA) have been efficiently used to achieve optimal or approximately optimal solutions without requiring prior knowledge about the data set to be clustered [3][4][5][6] .In a previous work GA and ACO algorithms were refined using local search in order to improve the clustering accuracy. Here, GA and ACO algorithms and their refinements through local search, and PSO and FA were developed to solve clustering problem. This document present them and their consequences data confirm that these metaheuristics are effectual to deal with the problem. The purpose function aims at minimize the square root of the sum of the squares of the differences between each object and its respective center. In this study we have developed methods like GA, ACO, PSO and FA that selects the best cluster centroid value than the existing clustering methods.

II. BACKGROUND STUDY

Clustering is disturbed with group jointly objects that are comparable to every other and dissimilar to the objects belong to other clusters. Clustering technique explore similarities between patterns and group similar pattern into categories or groups. In numerous fields there is obvious remuneration to be had from grouping jointly similar objects. For instance, in a medical application we strength wish to discover clusters of patients with similar symptoms. Grouping objects into category is a quite ordinary action and it has been intensify due to the big number of information that is currently available [1][2].

Evolutionary computation (EC) is a paradigm in the artificial intelligence realm that aims at benefiting from collective phenomena in adaptive populations of problem solvers utilizing the iterative progress comprising growth ,development, reproduction, selection, and survival as seen in a population . EAs employ this powerful design philosophy to find solutions to hard problems. EAs are cost based optimization algorithms the members of the EA relations split a great number of features in ordinary. They are all population-based stochastic investigate algorithms performing with best-to-survive criterion. Each algorithm commence by create an initial populace of possible solution, and evolve iteratively from generation to generation towards a best solution. In consecutive iterations of the algorithm, fitness-based assortment takes position within the populace of solutions. Improved solutions are preferentially chosen for continued existence into the next production of solution.

K-means algorithm with the robustness of the GAs to find an internationally optimal partition for a data set into a precise number of clusters. The reason of the GKA is to reduce the total intracluster variance. Though, its crossover operator is luxurious, so a number of change were future for this algorithm as in the quick Genetic K-Means Algorithm [7] and in the Incremental Genetic K-means Algorithm [8]. They reach a global optimum and are faster than GKA. One more algorithm proposed to resolve the same problem is the GA-clustering algorithm [9]. Its clustering metric is the sum of the Euclidean distances of the points from their respective cluster centers, as defined in Equation (1). The Genetic Algorithm for Clustering (GAC) developed in this work is based on the GA-clustering algorithm [9] because it reaches satisfactory results to solve the clustering problem.

Shelokar et al. [6] proposed an ACO algorithm for data clustering. It mostly relies on pheromone trails to direct ants to group objects according to their similarity and on a restricted search that randomly try to get better the most excellent iteration resolution earlier than updating pheromone trails. Their local search is perform as follow by selection of cluster number of every object is altered with a predefined probability .The ACO algorithm proposed by Kao and Cheng [5], called Ant Colony Optimization for Clustering (ACOC), attempts to improve the Shelokar's algorithm by introducing the concept of dynamic cluster centers in the ant clustering process, and by taking into account pheromone trails and heuristic information collectively at every solution building step. The ACO for Clustering presented in this paper is based on the ACOC algorithm proposed by Kao and Cheng [5].

ACO is among the most successful swarm based algorithms proposed by Dorigo & Di Caro in 1999 [11] .It is a meta heuristic inspired by the foraging behavior of ants in the wild, and moreover, the phenomena known as stigmergy, term introduced by Grasse in 1959.Stigmergy refers to the indirect communication amongst a self-organizing emergent system via individuals modifying their local environment. The most interesting aspect of the collaborative behavior of several ant species is their ability to find shortest paths between the ants' nest and the food sources by tracing pheromone trails Then, ants choose the path to follow by a probabilistic decision biased by the amount of pheromone trail, the higher its attractiveness. Because ants in turn deposit pheromone on the path they are subsequent, this behavior results in a self regulation process leading to the formation of paths marked by high pheromone meditation. By modelling and simulating ant foraging behavior, brood sorting, nest building and self-assembling, etc. algorithms can be developed that could be used for complex, combinatorial optimization problems. Particle swarm optimization (PSO) is a computational intelligence oriented, stochastic, population-based global optimization technique proposed by Kennedy and Eberhart in 1995[10]. It is inspired by the social behavior of bird flocking searching for food. PSO has been expansively applied to many manufacturing optimization area owing to its exclusive penetrating mechanism, straightforward concept, computational efficiency and simple implementation.

Most of fireflies fashioned short and musical flash and have dissimilar flashing behavior. Fireflies make use of these flashes for communiqué and attract the probable prey. YANG used this behavior of fireflies and introduces Firefly Algorithm in 2008 [11]. It is based on the supposition that solution of an optimization problem can be made known as a firefly which glow proportionally to its excellence in a measured problem location. Accordingly, each brighter firefly attracts its associates, which make the investigate freedom living being explored professionally. Yang used the FA for nonlinear design problems [12] and multimodal optimization problems [13] and showed the efficiency of the FA for finding global optima in two dimensional environments.

III METAHEURISTICS GA, PSO, ACO AND FA BASED METHODS

A. *GENETIC ALGORITHM (GA)*

Genetic algorithms are stimulated by the theory of natural selection and genetic evolution and they have been

ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 2, Issue 9, September 2013

productively applied to the optimization of complex processes. From an initial populace, basic operators are practical consisting of selection, crossover and mutation. The operators evolve the population generation to generation. All the way through the selection operator additional copy of individuals with the best fitness values are automatically allocated. The crossover operator combines part of two parent solutions to create a new solution. The mutation operator modifies at random the explanation fashioned by crossover. The descendent populace fashioned from the selection, crossover and mutation replace the parent population. There are a variety of techniques of replacement, for instance, elitism. The hybridization is a tremendously effectual way to augment presentation and competence of GAs. The majority ordinary form of hybridization is to integrate into GAs a method of local investigate as a important part in the progression, and also to include the field specific information in the search process. These types of methods with hybridization are called Memetic Algorithms (MA). The local search can be characterized as a local modification within a investigate space. Consequently, MA is connected to the intellectual evolution as individuals are adapted to meet the needs of the difficulty On the other hand, GA is based on the organic evolution of individuals, in such a way that the children will come into many skill and individuality present in their progenitors.GA-clustering algorithm since it reach satisfactory penalty to determine the cluster problem and it also model the problem as we describe at the establishment. The problem is represented by chromosomes consisting of an active array of c cluster come together centers. In an n-dimensional space, each middle is collected of n coordinates (n genes).The data structure includes, in addition to the centers, a reference to the data of those clusters, although they do not participate in the evolutionary procedure. The Memetic Algorithm for Clustering (MAC) is a version of the GAC with local search. The local search technique implement in the MAC is the First growth this is an alteration heuristic, a kind of Hill Climbing, which stop the development of the neighborhood when a improved neighbor is found.

1. Begin
2. Initialize the first generation data from the cluster p(t) and compute fitness
3. Initialize generation count to zero, N to 0
4. Perform genetic operation, before that from N individual of the best centroid vale select p individuals with fitness value and p(t-1)
5. Perform crossover operation with best cluster center value and increment the count, then Mutation is performed and finally Fitness function is calculated.
6. Sort the obtained values and Copy the first ten values of the present generation to the next generation
7. Go to step 3 if generation count is less than maxgen
8. End process with best centroid values and clustered data
9. End.

B. ***ANT COLONY ALGORITHM(ACO)***

ACO metaheuristic is an instance of an artificial swarm intelligence which is inspired by the collective behavior of communal insects. In the ACO algorithm, a reproduction ant simulate the pheromone trail subsequent the behavior of real ants to discover the shortest route between a food source and their nest. Every artificial ant collects the necessary information about the problem, stochastically make its own decision, and construct solution in a stepwise method. The behavior with the aim of emerge is a collection of comparatively "not intelligent" ants that interrelate throughout straightforward rules and energetically self-organize maintain their position approximately the straight trails. The pheromone script a trail, on behalf of a answer for a problem, which will be positively greater than before to develop into more good-looking in following iterations. So, the pheromone attentiveness indicate how helpful was an answer serving as a history of the best ants' previous movement. In addition the pheromone attentiveness, ants are able to use heuristic purpose values that usually point toward an open authority in the direction of more practical limited information.

A. In the ACOC, the solution space is represented by an object-cluster matrix containing o rows (objects) and c columns (clusters). Ants can stay in only one of the c clusters for each object. A vector (S) of size o is used to represent each solution built by ants. Each element of the vector tends to one of the o objects and it's accredited worth represent the cluster number assign to it. Each ant move on or after one node to other, deposit pheromone on nodes, and construct a solution in a stepwise manner. At every step, an ant selects an ungrouped object and adds it to its incomplete solution by bearing in mind both pheromone strength and heuristic information. Nodes with higher pheromone and heuristic principles would be added probable to be chosen by ants. The heuristic information indicates the attractiveness of assigning a data object to a particular cluster. It is obtained by calculating the mutual of the Euclidean distance between the data object to be grouped and each cluster center .Each ant carries a centers matrix (C_k) and updates it right after each clustering step

1. Initialize the pheromone matrix to minute values (τ);
2. Initialize all with cluster centers matrix (C_k) and the weight matrix (W_k) that associates each object with a center.
3. Select an object *i*: each ant selects an object *i*;
4. Select a cluster *j*: to determine *j* for a selected object *i*, two functions exploitation and exploration depending on the result

$$j = \begin{cases} \arg max_{j \epsilon N_i} \{[\tau(i,j)]^{\alpha} [\eta(i,j)]^{\beta} & if\ q \leq q_0 \\ S\ otherwise \end{cases}$$

(a) Exploitation allow ants to go in a greedy way to a node whose manufactured goods of pheromone level and heuristic value is the uppermost

(b) Exploration allows probability to candidate nodes, and then gives permission an ant choose one of them in a stochastic method according to

$$P^k(i,j) = \frac{[\tau(i,j)]^{\alpha} [\eta(i,j)]^{\beta}}{\sum_{j=1}^{g} 1\, [\tau(i,j)]^{\alpha} [\eta(i,j)]^{\beta}}$$

The more talented a node is, the higher its probability;
5. Update ants' matrices: update weight matrix (W_k) and cluster centers matrix;
6. Make sure each ant's solution: if the ant's solution vector (S_k) is complete, then go to step 7, otherwise, go back to step 3;
7. Calculate the objective function value of each ant .After that, rank the solutions of ants in the ascending order of objective function. The best solution is called iteration-best solution and the better one will be the new best-so-far solution;
8. Update pheromone trails: the global updating rule is applied, and only the elitist ants are allowed to add pheromone at the end of each iteration.
9. Check termination condition: if the number of iterations exceeds the predefined maximum iteration number, then it is stopped and the best-so-far solution is returned. Otherwise, go to step 2.

C. *PARTICLE SWARM OPTIMIZATION(PSO)*

PSO with biography Populations are organized according to some position of the data points in the cluster and select best cluster centroid to group the data. After selecting, then study their centroid values based on some predefined criteria and generates appropriate centroid values. Each particles move towards j is in i's neighborhood, i is also in j's. Each particle communicates with some other particles and is exaggerated by the best centroid point found by any member of its current centroid value p_i. The vector pi for that best neighbor, which we will denote with p_g. Initialize the particle's location best known position to its initial position: $p_i \leftarrow x_i$., then correspondingly update the particles or location of the centroid value position and their velocity to know the best global position or of the best centroid value to group the data .In PSO algorithm repeat the steps Until a termination criterion is met .Finally after finding the global best position only considers as best value of cluster centroid .

1. Initialize a population array of particles with random positions and velocities on D dimensions in the search space.
2. loop
3. For each particle, evaluate the desired optimization fitness function in D variables.
4. Compare particle's fitness evaluation with its $pbest_i$. If current value is better than$pbest_i$, then set $pbest_i$ equal to the current value, and Initialize the particle's location best known position to its initial position: $p_i \leftarrow x_i$
5. Identify the particle in the neighborhood with the best success and assign its index to the variable g.
6. If $(f(p_i) < f(g))$ update the swarm's best known position: $g \leftarrow p_i$
7. Initialize the particle's velocity: $v_i \sim U(-|b_{upv}-b_{lov}|, |b_{upv}-b_{lov}|)$
8. Change the velocity and position of the particle according to the following equation
9. Until a termination criterion is met ,repeat:
10. For each particle$(i = 1, ..., S)$
11. do
12. For each dimension $d = 1, ..., n$
13. do
14. Pick random numbers $r_p, r_g \sim U(0,1)$
15. Update the particle's velocity $v_{i,d} \leftarrow \omega\, v_{i,d} + \varphi_p r_p (p_{i,d}-x_{i,d}) + \varphi_g r_g (g_d-x_{i,d})$
16. Update the particle's position: $x_i \leftarrow x_i + v_i$
17. If $(f(x_i) < f(p_i))$ do:
18. Update the particle's best known position: $p_i \leftarrow x_i$
19. If $(f(p_i) < f(g))$ update the swarm's best known position: $g \leftarrow p_i$
20. Now g holds the best found solution.
21. end loop

D. *FIREFLY ALGORITHM (FA)*

In this proposed based firefly algorithm, the objective function (f(x)) of a given is based on difference in the cluster centroid value of the one data point to other data point with the best data point are final chosen. It helps the fireflies to travel towards best location of one data point in the cluster data to the other cluster data point and new attractive locations in order to obtain optimal data point to find the best centroid value in the data to group them into similar data points with cancer classification of without degradation of results of cancer diagnosis. After the evaluation of the initial population the firefly algorithm enters its main loop, which represents the greatest number of generation of the iterations for each firefly to find the best location of the centroid values and group the best centroid values into similar clusters. For each production the firefly with the greatest light concentration $(I_j > I_i)$ is chosen as the potential optimal solution. Each and every one firefly is characterized by their light intensity related with the objective function. Each firefly vary attractiveness with distance r via $\exp(-\lambda r)$; Vary attractiveness of the cluster datapoints update the best centroid value at each and every time when the new data points added or randomly other data points are chosen The population of n fireflies generates n solutions. Each firefly is changing its location iteratively. Finally Rank the fireflies and find the current best cluster centroid to group similar based data for cancer classification;

1. Initialize the objective function $f(x_i)$, $x = (x_1, x_2, \ldots\ldots, x_d)$
2. Generate an initial population for suitability index with fireflies $X_i\,(i = 1,2, \ldots\ldots, n)$
3. Set Max number of iterations=Max generations, Set t=1
4. Devise light intensity I so that it is linked with $f(x_i)$,
5. Define absorption coefficient λ
6. While (t<Maxgeneration)
7. For i=1: n (for all n fireflies)
8. For j=1: n (n fireflies)
9. If$(I_j > I_i)$
10. Move firefly i towards j;
11. $x_i^{t+1} = x_i^t + \beta \exp[-\gamma r_{ij}^2]$
12. End if

13. Vary attractiveness with distance r via $\exp(-\lambda r)$ update the cluster centriods
14. $r_{ij} = \|x_i - x_j\|_2 = \sqrt{\sum_{k=1}^{d}(x_{i,k} - x_{j,k})^2}$
15. Evaluate new solutions and update light intensity for each firefly;
16. End for j
17. End for i
18. Rank the fireflies and find the current best cluster data points;
19. End while
20. Post-processing the results and visualization of the location;
21. End process

IV.RESULT & DISCUSSION

In this section we measure the performance of the system in terms of the precision, recall, F-measure with technique medical dataset .Measuring these parameters show the results of the accuracy in terms of how clustering performance is improved.

A. *Precision*

Precision value is calculated is based on the retrieval of information at true positive (TP) prediction, false positive (FP).In privacy preservation data precision is calculated the percentage of preserved data results returned that are relevant.

$$Precision = TP/(TP+FP)$$

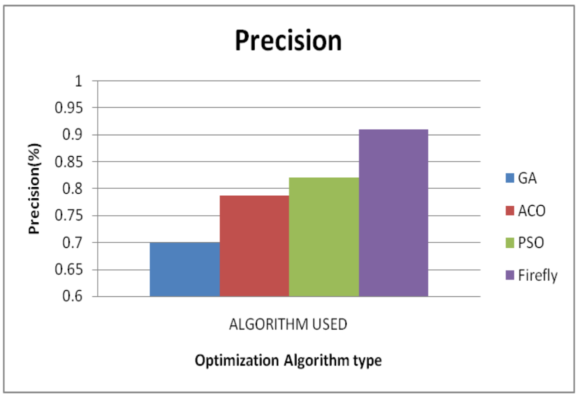

Fig.1 Precision Vs methods

From the above Fig.1 we can infer that Firefly algorithm is having the highest precision of 0.91 than the 0.82 of PSO, 0.787 of ACO and 0.7 of GA.

B. **Recall**

Recall value is calculated is based on the retrieval of information at true positive (TP) prediction, false negative (FP). In privacy preservation approach the data precision is calculated with percentage of positive results returned that are recall in this context is also referred to as the True Positive Rate (TP). Recall is the fraction of relevant instances that are retrieved.

$$Recall = TP/(TP+FN)$$

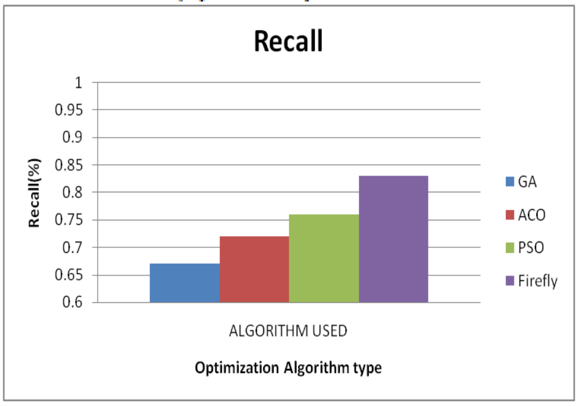

Fig.2 Recall Vs methods

From the above Fig.2 we can infer that Firefly algorithm is having the highest recall rate of 0.83 than the 0.76 of PSO, 0.72 of ACO and 0.67 of GA.

C. F-Measure

F-measure is a measure of a accuracy of test. It considers both the precision p and the recall r of the test to compute the score: p is the number of correct results divided by the number of all returned results and r is the number of correct results divided by the number of results that should have been returned. The score of F-measure can be interpreted as a weighted average of the precision and recall and usually a F1 score reaches its best value at 1 and worst score at 0.

$$Fmeasure = 2.Precision.recall/(precision + recall)$$

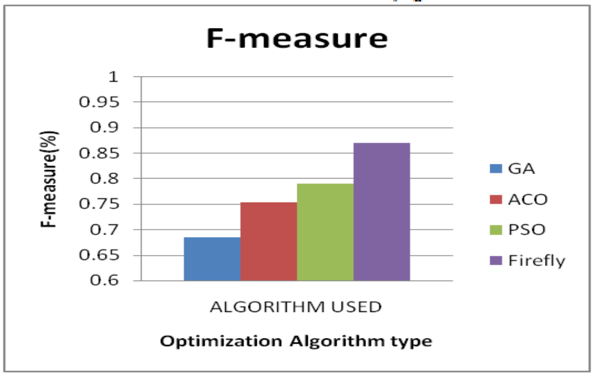

Fig.3 F-measure Vs methods

From the above Fig.3 we can infer that Firefly algorithm is having the highest F-measure of 0.87 than 0.79 of PSO, 0.7535 of ACO and 0.685 of GA.

V. CONCLUSION

In this work analyzed the performance of the GA, ACO, PSO and FA metaheuristics for solving data clustering problem in an experiment with breast cancer dataset data of UCI repository. From the evaluation experimental results we can infer that GA, ACO, PSO and FA are suitable metaheuristics to deal with this problem in the context of our experiment. The four algorithms presented in this paper called GAPF technique (GA, ACO, PSO and FA) were able to effectively discover clusters for the dataset used. These results also showed that, in our experiment, the Firefly optimization algorithm had better performance than all other optimization algorithms wither higher precision, recall and F-measure values. PSO is the second best algorithm as it had better performance than ACO and GA. However, ACO and GA achieved better fitness. On which, ACO is better than GA as it performed fewer objective function evaluations than GA. This is mainly due to two important characteristics of ACO. Firstly, it uses two information to guide the ants during the search of solutions, a history of the best ants previous movement (pheromone concentration) and an explicit influence toward more useful local information (heuristic function). Secondly, besides an exploration transition rule, it allows the ants to move in a greedy/deterministic manner to a node whose product of pheromone and heuristic value is the highest, i.e., when selecting a cluster to an object only the most similar one is generally chosen. Other observed behavior was that GA presented more variability on the results than the other algorithms resulted more inconsistent than all other algorithm. This behavior can be due to its crossover and mutation operators and need to be better analyzed in future works. ACO is also more elitist than GA once only the best ant adds pheromone at the pheromone matrix.

REFERENCES

[1] M. Bramer. “Principles of Data Mining”. Springer, 2007.

[2] J. Han and M. Kamber. “Data mining: Concepts and Techniques”. *Morgan Kaufmann Publisher, San Francisco*, 2001.

[3] R. H. Sheikh, M. M. Raghuwanshi, and A. N. Jaiswal. “Genetic Algorithm Based Clustering: A Survey”. *Proceedings of the 2008 First International Conference on Emerging Trends in Engineering and Technology (ICETET),* pp. 314-319, 2008.

[4] G. Garai and B. B. Chaudhuri. “A Novel Genetic Algorithm for Automatic Clustering”. *Pattern Recognition Letters*, vol. 25, n.2, pp.173-187, 2004.

[5] Y. Kao and K. Cheng. “An ACO-Based Clustering Algorithm”. *Lecture Notes in Computer Science*, vol. 4150, pp. 340-347, Springer, 2006.

[6] P. S. Shelokar, V. K. Jayaraman and B. D. Kulkarni. “An Ant Colony Approach for Clustering”. *Analytical Chimica Acta* 509, pp. 187–195, 2004.

[7] Y. Lu, S. Lu, and F. Fotouhi. “FGKA: A Fast Genetic K-means Clustering Algorithm”. *Proceedings of ACM SIG Symposium on Applied Computing (SAC’04), Nicosia, Cyprus*, pp. 622-623, 2004.

[8] Y. Lu, S. Lu, F. Fotouhi, Y. Deng, and S. J. Brown. “Incremental Genetic K-means Algorithm and its Application in Gene Expression Data Analysis”, *BMC Bioinformatics*, vol. 5, n. 172, 2004.

[9] U. Maulik and S. Bandyopadhyay. “Genetic Algorithm Based Clustering Technique”. *Pattern Recognition,* vol. 33, n. 9, pp. 1455-1465, 2000.

[10] Kennedy, J.; Eberhart, R. . "Particle Swarm Optimization". *Proceedings of IEEE International Conference on Neural Networks*. **IV**. pp. 1942–1948,1995.

[11] Dorigo, M., Maniezzo, V., & Colorni, A.,” Ant System: Optimization by a colony of cooperating agents,” *IEEE Transactions on Systems, Man, and Cybernetics – Part B*, 26, 29–41,1996.

[12] X. S. Yang, “Firefly algorithm، stochastic Test Functions and Design ptimization”. *Int. J. bio-inspired computation* .2010.

[13] X. S. Yang, “Firefly algorithm for multimodal optimization.” In: Stochastic Algorithms: foundations and applications ,”SAGA ‹*lecture notes in computer sciences*, pp. 169-178, 2009.

[14] D. W. van der Merwe and A. P. Engelbrecht, ‘‘Data Clustering Using Particle Swarm Optimization’’, *in the 2003 Congress on Evolutionary Computation*, Vol. 1, pp. 215-220, December 2003.