Proceedings of the 2007 IEEE Swarm Intelligence Symposium (SIS 2007)

MULTIPLE COOPERATING SWARMS FOR DATA CLUSTERING

Abbas Ahmadi †, *Fakhri Karray* ‡, *Mohamed Kamel* ‡

Pattern Analysis and Machine Intelligence Lab,
† Systems Design Engineering Department, ‡ Electrical and Computer Engineering Department,
200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1

***Abstract-* A new clustering technique by the use of multiple swarms is proposed. The proposed technique mimics the behavior of biological swarms which explore food situated in several places. We model the clustering problem using particle swarm optimization (PSO) approach. The proposed method considers multiple cooperating swarms to find centers of clusters. By assigning a portion of the solution space to each swarm, the exploration ability to find the solution is enhanced. Moreover, the cooperation among swarms increases the between-class distance. The proposed method outperforms k-means clustering as well as conventional PSO-based clustering techniques.**

***Index Terms*—** Multiple swarms, particle swarm optimization(PSO), clustering

1. INTRODUCTION

Particle Swarm Optimization is a search method that imitates the swarming behavior of flocks of birds, schools of fish, and swarms of bees [1, 2], first introduced by Kennedy and Eberhart [3, 4]. Like Genetic Algorithms(GAs), it employs a population of individuals known as particles to solve the optimization problem. As compared to GAs, a swarm is similar to a population, whereas a particle behaves the same as an individual. PSO tries to optimize an objective function f, called fitness function. It starts from an initial population and explores the solution space through a number of iterations to reach a near optimal solution. Each particle is characterized by a position-vector x_i and velocity-vector v_i. There is also a best position for each particle, known as best personal, denoted by x_i^{bp}. Furthermore, there is a best position-vector among the swarm, also called global best, denoted by x_i^*. A new velocity and position of each particle is obtained by the use of the following equations, respectively

$$v_i(t+1) = wv_i(t)+c_1r_1(x_i^{bp}(t)-x_i(t))+c_2r_2(x^*(t)-x_i(t)), \tag{1}$$

$$x_i(t+1) = x_i(t) + v_i(t+1), \tag{2}$$

where w is inertia weight to control the impact of the previous history of velocities on the current one. Also, c_1 and c_2 are positive constants known as self-recognition component and social component, respectively. Moreover, r_1 and r_2 are samples of random variables uniformly distributed in the interval $[0, 1]$; i.e., $r_1, r_2 \sim U(0, 1)$. As can be seen from (1), to produce a new position each particle follows two best values, which are best personal and global best of swarm obtained so far. In the case that fitness function is minimized, the best personal position of a particle i at iteration t is updated as follows:

$$x_i^{bp}(t+1) = \begin{cases} x_i^{bp}(t) & \text{if } f(x_i(t+1)) \geq f(x_i^{bp}(t)) \\ x_i(t+1) & \text{otherwise.} \end{cases} \tag{3}$$

The best particle of the swarm is also updated using the following equation:

$$x^*(t+1) = arg \min_{x_i(t)} \{f(x_1(t)), ..., f(x_n(t)), f(x_{n+1}(t))\}, \tag{4}$$

where $x_{n+1}(t)$ is the global best at previous iteration. The initial velocities can be set to zero; i.e.,

$$v_i(0) = 0, \quad i = 1, ..., n. \tag{5}$$

One can also initiate velocities by generating random values. The best personal for each particle is initialized as follows:

$$x_i^{bp}(0) = x_i(0), \;\; i = 1, ..., n. \tag{6}$$

There are several methods to terminate PSO procedure, such as maximum number of iterations, number of iterations with no improvement, and minimum objective function criterion [5]. We use the first method as our termination criterion.

Here, we explain motivations to apply particle swarm optimization for clustering purposes. We enumerate two main categories as motivations: biological and computational.

1. ***Biological motivations***

   Historically, the biological behavior of swarms was the main motivation behind particle swarm optimization [1, 5]. PSO founders mimicked swarming behavior of flocks of birds, schools of fish or swarms of bees whose goal

1-4244-0708-7/07/$20.00 ©2007 IEEE

are usually finding food. In canonical PSO, food is located in a single point and the swarm tend to reach that point. However, there are occasions in which there are more than one possible point to get food; for instance in the case of bees, usually there exist more than one possible bunch of flowers. In other words, there are multiple swarms, each of which is looking for a bunch of flowers. During search, each swarm cooperates with other swarms to find a proper place -that is a bunch of flowers- far enough from the others. Moreover, within each swarm, particles prefer to go dense places where the compactness of flowers is as high as possible.

2. ***Computational motivations***

Computational issues have also stimulated employing particle swarm optimization for clustering. In the following, we enumerate these motivations:

- The PSO algorithm performs a global search of solution space, whereas most other clustering techniques perform a local search [6]. In the local search, the solution obtained is located in the vicinity of the previous solution. For example, the $k-$means clustering method applies the randomly generated points as the initial centers of clusters and updates the position of the centers at every iteration. This may cause the algorithm to converge to suboptimal solutions. At the same time, PSO is less sensitive to the effect of the initial conditions due to its population-based nature. Therefore, it is more probable to find near optimal solutions. Xiang et al. have employed hybrid PSO and self-organizing maps to construct a scheme for gene clustering [7]. First, the weights are trained using self-organizing maps. These weights are then optimized applying PSO. Cui et al. have proposed a new method for document clustering [8]. In their work, PSO tries to find optimal centers of clusters in the solution space based on the average distance of documents from their corresponding centers, which is used as the fitness function to evaluate the solution represented by each particle. Omran et al. have applied PSO for image clustering [6, 9]. Their proposed method is similar to the previous work and the main difference is the way they define fitness function. They tend to cluster images such that intra-cluster distance and quantization error minimized while the distance between clusters maximized.
- Particle swarm optimization has been used to solve multi-objective optimization problems [2]. In terms of optimization, clustering can be a multi-objective problem. On one hand, we desire to have as compact clusters as possible. On the other hand, we prefer to have as separate as possible clusters. Conventional clustering techniques such as $k-$means usually consider only the former criterion, whereas the PSO-based clustering technique can deal with multiple objectives.
- Both multiple and cooperative swarms have been also introduced to solve optimization problems [10, 11, 12]. Bergh and Engelbrecht have used cooperative multiple swarms to solve optimization problems [10]. Their proposed method performs better than single swram in high dimensions due to the exponential increase in the volume of the search space as the dimension of the problem increases. This idea is valid for clustering problems as well. When the dimensionality of the data is high and the number of clusters is large, the ability of a single swarm is not sufficient to search all the solution space. Instead, multiple swarms cooperating together can be employed to obtain cluster centers effectively.

In the following section, we present conventional single swarm-based clustering. Next, we introduce our proposed clustering technique based on multiple swarms. In section 4, we present experimental results. We finally draw conclusions and give future research directions.

2. SINGLE SWARM-BASED CLUSTERING TECHNIQUE

PSO is a search technique which is mainly introduced to deal with optimization problems [1, 2]. Due to its abilities, it has been applied in other applications such as classification and clustering. To use particle swarm optimization as a clustering technique, one should model the clustering task as an optimization problem. The goal of such a model is to obtain centers of clusters so that some objective function is optimized. Assume $\mathcal{Y}$ is a set of data points intended to be clustered into K separate clusters given by

$$\mathcal{Y} = \{y^1, y^2, ..., y^K\}, \qquad (7)$$

where y^k indicates k^{th} cluster's data. Also, suppose n_k is the number of data samples in cluster k and m^k denotes the center of cluster k. Moreover, particle i is modelled by $x_i = (m^1, ..., m^K)_i$. We next provide the mathematical model of the single swarm-based clustering in terms of the optimization problem. To model the clustering problem as an optimization problem, it is required to formulate objective functions as well as constraints. There is no constraint but having points selected from the domain of data set or solution space. The objective function can be modelled by means of the compactness and separation measures. These measures are usually used to evaluate the performance of clustering techniques.

1. **Compactness:** This measure is related to within-cluster distance. Considering $m^1, ..., m^K$ as the centers of clusters $1, ..., K$ the compactness of clusters is obtained by

$$F_c(m^1, ..., m^K) = \frac{1}{K}[\sum_{k=1}^{K}(\frac{1}{n_k}\sum_{j=1}^{n_k} dist(m^k, y_j^k))], \tag{8}$$

where $dist(.)$ stands for the Euclidean distance between cluster center, m^k, and the data point y_j^k [13]. The goal is to minimize this function as much as possible.

2. **Separation:** This criterion shows how far the clusters are from each other. The clusters' separation measure can be formulated by

$$F_s(m^1, ..., m^K) = \frac{1}{K(K-1)}[\sum_{j=1}^{K}\sum_{k=j+1}^{K} dist(m^j, m^k)]. \tag{9}$$

This measure computes the accumulative distance of cluster centers from each other. Clustering techniques aim to maximize this criterion [13].

Having the inter-cluster and intra-cluster distance measures defined, we can now construct the objective function for the problem. Here, we deal with a multi-objective optimization problem containing two different functions, called $F_c(\cdot)$ and $F_s(\cdot)$. The former objective function should be minimized, whereas the last one needs to be maximized. By knowing that $max\ f(x)$ is equivalent to $min\ (-f(x))$, the weighted sum of the objective functions can be given by

$$F(m^1, ..., m^K) = w_1 F_c(m^1, ..., m^K) - w_2 F_s(m^1, ..., m^K), \tag{10}$$

where $w_1 + w_2 = 1$. We can now use PSO to obtain the solution of the problem as we formulated the required objective function. The search starts from an initial population in the feasible solution space and proceeds to find a near optimal solution.

After formulating the clustering problem by the notion of single swarm, we introduce the multiple swarms-based clustering technique in the following section. In our proposed method within each swarm, all particles attempt to find the best point as cluster center. There is also an information exchange or cooperation between swarms and each swarm obtains its corresponding center while interacting with others.

3. MULTIPLE SWARMS-BASED CLUSTERING

In this section, we propose a new technique for clustering problem using multiple swarms. We consider the following assumptions:

- The number of swarms is as many as that of clusters. That is, each swarm corresponds to a cluster.
- Each swarm is responsible to find its corresponding cluster's center.
- Particles of each swarm are candidates for the corresponding cluster's center.

The whole procedure to reach an optimal solution in the proposed method is done through two main phases, including initialization and exploration as discussed next.

1. **Initialization phase**

The search first starts from random points in the solution space. In the beginning of this phase, there is overlap between swarms, whereas at the end each swarm will deal with a part of the solution space. The situation of swarms at the beginning and end of initialization phases is shown in Fig. 1.

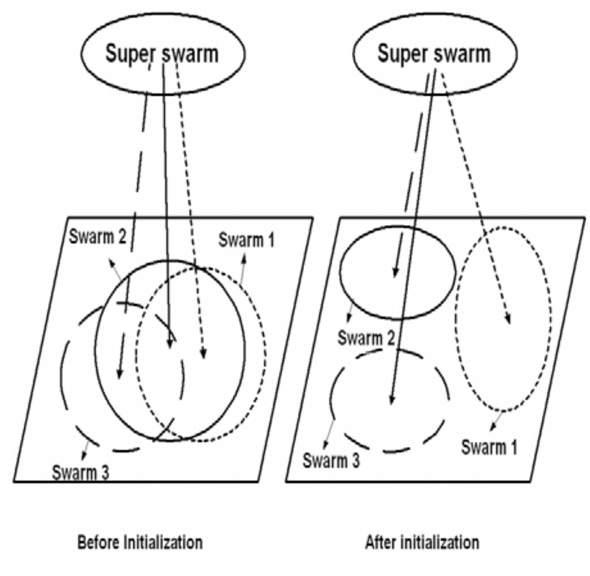

Fig. 1. Situation of swarms at the beginning and end of initialization phases

Each swarm is responsible to search a part of solution space, called exploration space of the swarm. The exploration space of each swarm is characterized by two parameters as follows:

- center of swarm region (z^k),
- width of swarm region (r^k).

The first parameter denotes the center of the exploration space and the second one shows its width. The main goal of the initialization phase is to determine these parameters for all swarms.

To perform the initialization phase, we have used another swarm, called super swarm. Super swarm obeys the single swarm-based clustering technique to direct swarms to the dense places as PSO procedure can scape from local optima. In this phase, each swarm receives information from only the super swarm. First, the super swarm searches for the center of the swarm regions. This information is then supplied to all swarms. Finally, the swarms try to move toward these centers. By performing one iteration within each swarm, its corresponding width is updated. To obtain the widths, we use eigen decomposition theorem. Let's assume λ^k denotes the square root of the biggest eigen value of data belonging to swarm k. The width of swarm region k is then computed by

$$r^k = \alpha\lambda^k, \tag{11}$$

where α is a positive constant. The appropriate value for α is selected such that the distribution of swarms would look like the scheme c of Fig. 2.

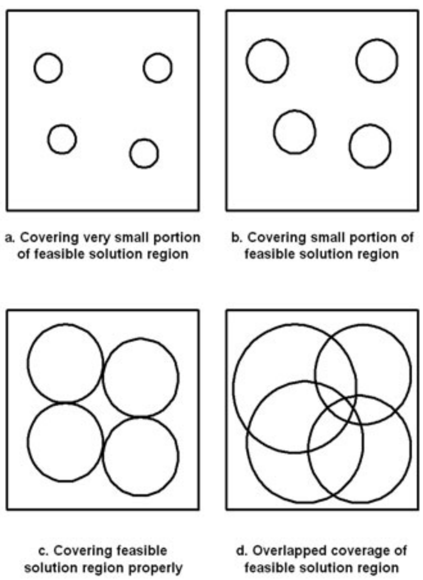

Fig. 2. Width of swarm regions

After updating the width of the swarm regions, a new iteration begins. Again, the super swarm updates the centers of the swarm regions. These new centers are fed to the swarms. The initialization phase ends when a maximum number of iterations is achieved.

2. **Exploration phase**

After initializing swarms, each swarm explores for the best solution as cluster center within its corresponding region. In this phase, there is no super swarm, but rather information exchange between swarms. Hence, there is a cooperation between swarms to find the final solution. Each swarm knows the global best of the other swarms.

This phase contains a number of iterations to converge to a near optimal solution. Each iteration comprises two main steps: search and make decision. In the first step, search within each swarm region proceed. In the next step, it is revealed that whether the new solution is acceptable or not. We next provide more explanations about both steps.

- Search

In this step, search within each swarm region is done in a way that within-cluster distance is minimized, and at the same time the accumulated distance from other clusters is maximized. Within-cluster distance deals with compactness of the cluster. The compactness of cluster k given particle i as its center is obtained by

$$f_c(x_i^k) = \frac{1}{n_k}[\sum_{j=1}^{n_k} dist(x_i^k, y_j^k)], \tag{12}$$

where x_i^k is particle i of swarm k. Also, $dist(.)$ stands for the Euclidean distance between cluster center, x_i^k, and cluster's data points, y_j^k.

Distance from other clusters shows how far that particular cluster is from other clusters. This distance for particle i of swarm k can be formulated as follows:

$$f_s(x_i^k) = \frac{1}{K-1}[\sum_{l=1}^{K} dist(x_i^k, m^l)]. \tag{13}$$

Thus, the objective function for particle i of swarm k, x_i^k, is given by

$$f(x_i^k) = w_1 f_c(x_i^k) - w_2 f_s(x_i^k), \tag{14}$$

where $w_1 + w_2 = 1$. After defining the objective function of the problem, the mathematical model of the clustering task -in terms of optimization problem by the use of multiple swarms- can be constructed. In search step within each swarm, particles attempt to minimize the following optimization problem:

$$\begin{aligned} \min \ & f(x_i^k) = w_1 f_c(x_i^k) - w_2 f_s(x_i^k) \\ & s.t. : \|x_i^k - z^k\| \leq r^k \end{aligned} \tag{15}$$

In this equation, the constraint forces particles of the swarm to search within its corresponding region.

Search using multiple swarms is performed in a serial scheme. It is begun in first swarm region and a new candidate for the cluster center is obtained using equation (15). Considering this new candidate, the next swarm searches for a new candidate to its corresponding cluster center. Similarly, this procedure is repeated for each of the following swarms to obtain new candidates for centers of all clusters.

- Make decision

  When search for all swarms is completed, it is necessary to decide on the new candidates for centers of clusters: accept or reject. If the amount of fitness function obtained by equation (10) for new candidates $(m^1, ..., m^K)$ is less than that of the former iteration, the new solution is accepted. Otherwise, it is not.

Having explained the procedure in full, we provide algorithm of the proposed method in Algorithm **1**.

Algorithm 1 Multiple swarms-based clustering

Phase 1: Initialization by the super swarm

- Determine swarms' center
- Determine swarms' width

Phase 2: Exploration

- Step 1: Search within each swarm
  - 1.1. Compute new positions of all particles of swarms.
  - 1.2. Obtain the fitness value of all particles using equation (14).
  - 1.3. Select that position which minimizes optimization problem (15) and denote it as new candidate for corresponding cluster center.
- Step 2: Make decision
  - 2.1. Calculate the fitness value of the new candidates for centers of clusters using equation (10).
  - 2.2. If the fitness value is smaller than the that of previous iteration, accept the new solution; otherwise, reject it.
  - 2.3. If termination criterion is achieved, stop; otherwise, go step 1.

---

4. EXPERIMENTAL RESULTS

In this section, we examine the performance of the proposed method and compare it with k-means as well as single swarm-based clustering.

We executed experiments on three sets of data, as follows:

- Gaussian data: a total of 800 samples drawn from four two-dimensional Gaussian classes with following distribution:

$$N(\mu = \begin{pmatrix} m_i \\ 0 \end{pmatrix}, \ \sum = \begin{bmatrix} 0.50 & 0.05 \\ 0.05 & 0.50 \end{bmatrix}), \qquad (16)$$

  where μ denotes the mean vector and $\sum$ is the covariance matrix; $m_1 = -3$, $m_2 = 0$, $m_3 = 3$ and $m_4 = 6$.

- Speech data: four phonemes /aa/, /ae/, /ay/, and /el/ from TIMIT database [14] have been considered. A total of 800 samples from those classes have been taken. We have used 12 mel-frequency cepstral coefficients (MFCCs)[15] and speech data sampled at 16 kHz.

- Zoo data: seven classes of animals from CMI machine learning repository have been selected. The dimensionality of data is 17 and there exist 101 objects.

Fig. 3 shows the performance of the proposed method over speech data for 80 iterations.

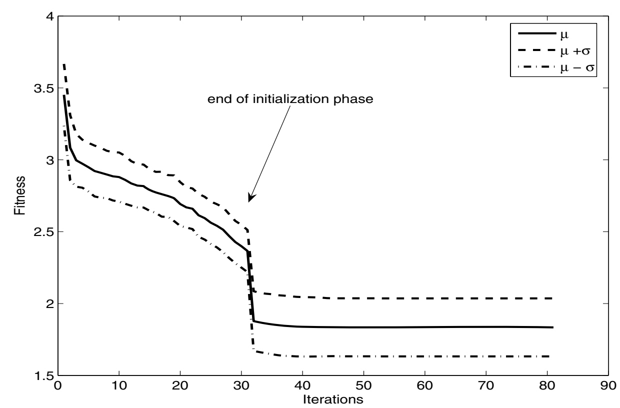

Fig. 3. Convergence of the proposed method in terms of fitness function: The values for fitness function are obtained by averaging over multiple runs(30 runs). The standard deviation (std) of fitness function is also shown by dash and dash-dot lines.

As shown in Fig. 3, initialization phase is completed after 30 iterations. Due to the cooperations among multiple swarms, a significant improvement is observed at the beginning of the exploration phase.

The parameters in the model are considered as $c_1 = 2$, $c_2 = 1$, $w_1 = 0.85$, $n = 10$. Moreover, we have considered $\alpha = 1.9$, $\alpha = 1$ and $\alpha = 2$ for Gaussian, speech and zoo data, respectively.

We have compared our proposed method with k-means clustering and single swarm-based clustering using Gaussian, speech and zoo data (Table 1, Table 2 and Table 3). The comparison is based on the compactness, separation and fitness value. The results are obtained by averaging over 30 different runs. For each value, we have also provided its associated standard deviation.

Table 1. COMPARISON FOR GAUSSIAN DATA

Method	Compactness	Separation	Fitness
K-means	532.92 ± 61.94	29.99 ± 0.52	-0.01 ± 0.08
Single swarm	582.08 ± 79.31	34.57 ± 2.88	-0.03 ± 0.1
Proposed	522.16 ± 37.29	31.03 ± 0.87	-0.04 ± 0.02

Table 2. COMPARISON FOR SPEECH DATA

Method	Compactness	Separation	Fitness
K-means	$2.41e003 \pm 22.6$	25.96 ± 0.16	2.59 ± 0.07
Single swarm	$3.5e003 \pm 227.7$	72.35 ± 8.56	1.89 ± 0.11
Proposed	$2.41e003 \pm 41.3$	34.32 ± 0.67	1.81 ± 0.06

Table 3. COMPARISON FOR ZOO DATA

Method	Compactness	Separation	Fitness
K-means	97.1511 ± 7.22	66.11 ± 5.2	-0.10 ± 0.13
Single swarm	132.53 ± 8.35	104.42 ± 14.7	-0.11 ± 0.103
Proposed	130.058 ± 8.86	87.77 ± 6.81	-0.17 ± 0.16

In Fig. 4, 5 and 6, we have compared the behavior of fitness function through 80 iterations for the proposed method, $k-$means and single swarm-based clustering over Gaussian, speech and zoo data, respectively.

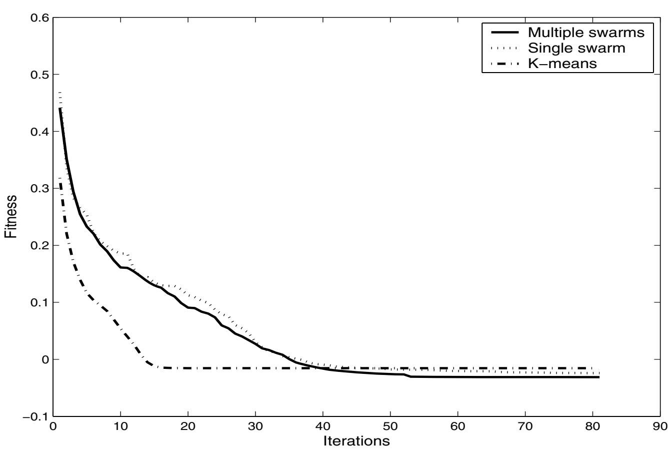

Fig. 4. Performance of the proposed multiple swarms-based, k-means and single swarm-based clustering in terms of fitness value (using Gaussian data)

As illustrated in Fig. 4, 5 and 6, $k-$means clustering technique converges quickly, but multiple swarms-based clustering technique can provide better solutions in terms of fitness function due to its strong and effective search ability. In

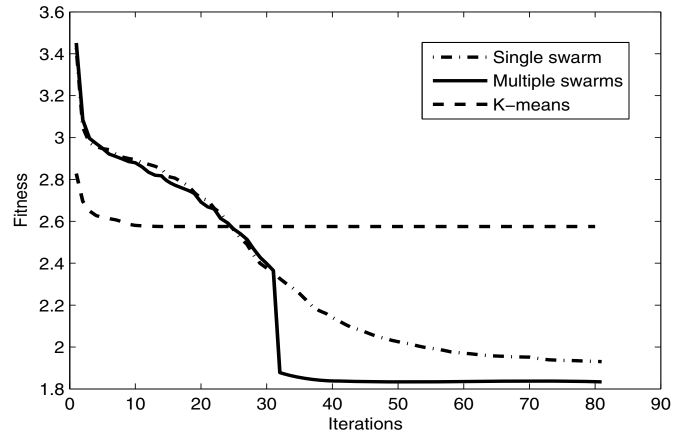

Fig. 5. Performance of the proposed multiple swarms-based, k-means and single swarm-based clustering in terms of fitness value (using speech data)

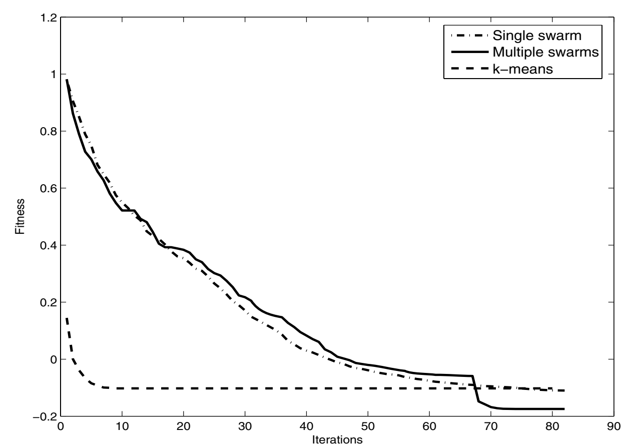

Fig. 6. Performance of the proposed multiple swarms-based, k-means and single swarm-based clustering in terms of fitness value (using zoo data)

terms of computational time, $k-$means provides better solutions. Meanwhile, multiple swarm-based technique outperforms single swarm-based technique as it designate a portion of search space to each swarm.

5. CONCLUSION

In this paper, we proposed a new clustering technique on the basis of multiple swarms. The idea originated from the biological behavior of multiple swarms which are looking for food located in several places. Clustering problems usually contain several objectives to be optimized. The proposed technique is capable of considering several objective functions simultaneously. It assigns a portion of the solution space to each swarm. This strategy boosts its exploration ability as each swarm deals with a part of solution space. Each swarm explores its own region while cooperating with other swarms. It knows the global best of other swarms and attempts to find a point whose accumulative distance from the other clusters' centers is maximum. Each swarm also tends to decrease within-class distance.

Our proposed method is applied for clustering three sets of data comprising Gaussian, speech and zoo data. The proposed method outperforms the other methods because of the following reasons:

- considering multiple objectives in the model,
- using multiple cooperating swarms,
- having strong search ability due to assigning the search space to multiple swarms.

There are other issues which influence the performance of the problem. Alternatives for fitness functions and other choices for distance measure are among those issues to be explored in future works.

6. REFERENCES

[1] James Kennedy; Russell C. Eberhart and Yuhui Shi. *Swarm intelligence*. Morgan Kaufman Publishers, 2001.

[2] A.P. Engelbrecht. *Fundamentals of Computational Swarm Intelligence*. John Wiley and Sons, first edition, 2005.

[3] J. Kennedy and R.C. Engelbrech. Particle swarm optimization. In *IEEE International Conference on Neural Networks*, volume 4, pages 1942–1948, 1995.

[4] R.C. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *6th Int. Symp. Micro Machine and Human Scince*, pages 39–43, Nagoya, Japan, 1995.

[5] Ajith Abraham; He Guo and Hongbo Liu. *Studies in Computational Intelligence*, chapter Swarm Intelligence: Foundations, Perspectives and Applications, Swarm Intelligent Systems. Springer Verlag Germany, 2006.

[6] Mahamed G. H. Omran ; Ayed Salman and Andries P. Engelbrecht. Dynamic clustering using particle swarm optimization with application in image segmentation. *Pattern Analysis and Applications*, 6:332 – 344, 2006.

[7] Xiang Xiao; E.R. Dow; R. Eberhart; Z.B. Miled; R.J. Oppelt. Gene clustering using self-organizing maps and particle swarm optimization. In *IEEE Proceedings of International Parallel Processing Symposium (IPDPS'03)*, page 10pp, 2003.

[8] X. Cui; T.E. Potok; P. Palathingal. Document clustering using particle swarm optimization. In *IEEE Proceedings of Swarm Intelligence Symposium, 2005 (SIS 2005)*, pages 185 – 191, 2005.

[9] M. Omran; A. P. Engelbrecht and A. Salman. Particle swarm optimization method for image clustering. *International Journal of Pattern Recognition and Artificial Intelligence*, 19(3):297321, 2005.

[10] F. van den Bergh and A.P. Engelbrecht. A cooperative approach to particle swarm optimization. *IEEE Transactions on Evolutionary Computing*, 8(3):225–239, June 2004.

[11] M. El-Abd and M. Kamel. Information exchange in multiple cooperating swarms. In *Proceedings of IEEE Swarm Intelligence Symposium (SIS 2005)*, pages 138 – 142, 2005.

[12] T. Blackwell and J. Branke. *Applications of Evolutionary Computing*, volume 3005 LNCS, chapter Multiswarm Optimization in Dynamic Environments, pages 488–599. Springer, 2004.

[13] Richard O. Duda; Peter E. Hart and David G. Stork. *Pattern Classification*. John Wiley & Sons, 2000.

[14] National Institute of Standards and Technology Speech Disc 1-1.1, NTIS Order No. PB91-5050651996. *Timit acoustic-phonetic continuous speech corpus*, October 1990.

[15] John Coleman. *Introducing speech and language processing*. Cambridge university press, 2005.

[16] F. Karray and C. de Silva. *Soft computing and intelligent systems design: Theory, Tools and Applications*. Addison Wesley Publishing, 2004.

[17] Fun Ye and Ching yi Chen. Alternative kpso-clustering algorithm. *Tamkang Journal of Science and Engineering*, 8(2):165–174, 2005.

[18] Ching yi Chen and Fun Ye. Particle swarm optimization algorithm and its application to clustering analysis. In *Proceeding of the 2004 IEEE International Conference on Networking, Sensing and Control*, pages 789–794, 2004.