International Journal of Research in Computer Science
eISSN 2249-8265 Volume 3 Issue 2 (2013) pp. 1-9
www.ijorcs.org, A Unit of White Globe Publications
doi: 10.7815/ijorcs. 32.2013.060

A PSO-BASED SUBTRACTIVE DATA CLUSTERING ALGORITHM

Mariam El-Tarabily[1], Rehab Abdel-Kader[2], Mahmoud Marie[3], Gamal Abdel-Azeem[4]

[1 2 4]*Electrical Engineering Department, Faculty of Engineering - Port-Said, Port-Said University, EGYPT*
E-mail: [1]mariammokhtar75@hotmail.com, [2]r.abdelkader@eng.psu.edu.eg, [4]gamalagag@hotmail.com

[3]*Computers and Systems Engineering Department, Faculty of Engineering, Al-Azhar University, Cairo, EGYPT*
E-mail: mahmoudim@hotmail.com

Abstract: *There is a tremendous proliferation in the amount of information available on the largest shared information source, the World Wide Web. Fast and high-quality clustering algorithms play an important role in helping users to effectively navigate, summarize, and organize the information. Recent studies have shown that partitional clustering algorithms such as the k-means algorithm are the most popular algorithms for clustering large datasets. The major problem with partitional clustering algorithms is that they are sensitive to the selection of the initial partitions and are prone to premature converge to local optima. Subtractive clustering is a fast, one-pass algorithm for estimating the number of clusters and cluster centers for any given set of data. The cluster estimates can be used to initialize iterative optimization-based clustering methods and model identification methods. In this paper, we present a hybrid Particle Swarm Optimization, Subtractive + (PSO) clustering algorithm that performs fast clustering. For comparison purpose, we applied the Subtractive + (PSO) clustering algorithm, PSO, and the Subtractive clustering algorithms on three different datasets. The results illustrate that the Subtractive + (PSO) clustering algorithm can generate the most compact clustering results as compared to other algorithms.*

Keywords: *Data Clustering, Subtractive Clustering, Particle Swarm Optimization, Subtractive Algorithm, Hybrid Algorithm.*

I. INTRODUCTION

Clustering is one of the most extensively studied research topics due to its numerous important applications in machine learning, image segmentation, information retrieval, and pattern recognition. Clustering involves dividing a set of objects into a specified number of clusters [14]. The motivation behind clustering a set of data is to find inherent structure in the data and expose this structure as a set of groups. The data objects within each group should exhibit a large degree of similarity while the similarity among different clusters should be minimized [3, 9, 18]. There are two major clustering techniques: "Partitioning" and "Hierarchical" [2, 9]. In hierarchical clustering, the output is a tree showing a sequence of clustering with each clustering being a partition of the data set. On the other hand, Partitioning clustering [1] algorithms partition the data set into a specified number of clusters. These algorithms try to minimize a certain criteria (e.g. a square error function) and can therefore be treated as optimization problems.

In recent years, it has been recognized that the partitional clustering technique is well suited for clustering large datasets due to their relatively low computational requirements. The time complexity of the partitioning technique is almost linear, which makes it a widely used technique. The best-known partitioning clustering algorithm is the K-means algorithm and its variants [10].

Subtractive clustering method, as proposed by Chiu [13], is a relatively simple and effective approach to approximate estimation of cluster centers on the basis of a density measure in which the data points are considered candidates for cluster centers. This method can obtain initial cluster centers that are required by more sophisticated clustering algorithms. It can also be used as quick stand-alone method for approximate clustering.

Particle Swarm Optimization (PSO) algorithm is a population based stochastic optimization technique that can be used to find an optimal, or near optimal, solution to a numerical and qualitative problem [4, 11, 17]. Several attempts were proposed in the literature to apply PSO to the data clustering problem [6, 18, 19, 20, 21]. The major drawback is that the number of cluster is initially unknown and the clustering result is sensitive to the selection of the initial cluster centroids and may converge to the local optima. Therefore, the initial selection of the cluster centroids decides the processing of PSO and the partition result of the dataset as well. The same initial cluster centroids in a dataset will always generate the same cluster results. However, if good initial clustering centroids can be obtained using any of the other techniques, the PSO

would work well in refining the clustering centroids to find the optimal clustering centers. The Subtractive clustering algorithm can be used to generate the number of clusters and a good initial cluster centroids for the PSO.

In this paper, we present a hybrid Subtractive + (PSO) clustering algorithm that performs fast clustering. Experimental results indicate that the Subtractive + (PSO) clustering algorithm can find the optimal solution after nearly 50 iterations in comparison with the ordinary PSO algorithm. The remainder of this paper is organized as follows: Section 2 provides the related works in data clustering using PSO. Section 3 provides a general overview of the data clustering problem and the basic PSO algorithm. The proposed hybrid Subtractive + (PSO) clustering algorithm is described in Section 4. Section 5 provides the detailed experimental setup and results for comparing the performance of the Subtractive + (PSO) clustering algorithm with the Subtractive algorithm, and PSO. The discussion of the experiment's results is also presented. Conclusions are drawn in Section 6.

II. RELATED WORK

The well-known partitioning algorithm is the K-means algorithm [2, 6, 7, 16] and its variants. The main drawback of the K-means algorithm is that the cluster result is sensitive to the selection of the initial cluster centroids and may converge to the local optimal and it generally requires a prior knowledge of the probable number of clusters for a data collection. In recent years scientists have proposed several approaches [3] inspired from the biological collective behaviors to solve the clustering problem, such as Genetic Algorithm (GA) [8], Particle Swarm Optimization (PSO), Ant clustering and Self-Organizing Maps (SOM) [9]. In [6] authors represented a hybrid PSO+K-means document clustering algorithm that performed fast document clustering. The results indicated that the PSO+K-means algorithm can generate the best results in just 50 iterations in comparison with the K-means algorithm and the PSO algorithm. Reference [24] proposed a Discrete PSO with crossover and mutation operators of Genetic Algorithm for document clustering. The proposed system markedly increased the success of the clustering problem, it tried to avoid the stagnation behavior of the particles, but it could not always avoid that behavior. In [26] authors investigated the application of the EPSO to cluster data vectors. The EPSO algorithm was compared against the PSO clustering algorithm which showed that the EPSO convergence is slower to lower quantization error, while the PSO convergence is faster to a large quantization error. Reference [27] presented a new approach to particle swarm optimization (PSO) using digital pheromones to coordinate swarms within an n-dimensional design space to improve the search efficiency and reliability. In [28] a hybrid fuzzy clustering method based on FCM and fuzzy PSO (FPSO) is proposed which make use of the merits of both algorithms. Experimental results show that the proposed method is efficient and can reveal encouraging results.

III. DATA CLUSTERING PROBLEM

In most clustering algorithms, the dataset to be clustered is represented as a set of vectors $X = \{x_1, x_2, \ldots, x_n\}$, where the vector x_i corresponds to a single object and is called the feature vector. The feature vector should include proper features to represent the object.

The similarity metric: Since similarity is fundamental to the definition of a cluster, a measure of the similarity between two data sets from the same feature space is essential to most clustering procedures. Because of the variety of feature types and scales, the distance measure must be chosen carefully. Over the years, two prominent ways have been proposed to compute the similarity between data m_p and m_j. The most popular metric for continuous features is the Euclidean distance, given by:

$$d_{(m_p,m_j)} = \sqrt{\sum_{k=1}^{d_m} (m_{pk} - m_{jk})^2 / d_m} \quad (1)$$

which is a special case of the Minkowski metric [5] , given by:

$$D_n(m_p, m_j) = \left(\sum_{i=1}^{dm} |m_{i.p} - m_{i,j}|^n \right)^{1/n} \quad (2)$$

where m_p and m_j are two data vectors; d_m denotes the dimension number of the vector space; m_{pk} and m_{jk} stand for the data m_p and m_j's weight values in dimension k.

The second commonly used similarity measure in clustering is the cosine correlation measure [15], given by:

$$\cos(m_p, m_j) = \frac{m_{pm_j}^t}{|m_p||m_{j|}} \quad (3)$$

where m_p^t, m_j denotes the dot-product of the two data vectors; |.| indicates the length of the vector. Both similarity metrics are widely used in clustering literatures.

A. Subtractive Clustering Algorithm

In Subtractive clustering data points are considered as candidates for the cluster centers [25]. In this method the computation complexity is linearly proportional to the number of data points and

independent of the dimension of the problem under consideration.

Consider a collection on n data points $\{x_1, \ldots, x_n\}$ in an M-dimensional space. Since each data point is a candidate for cluster centers, a density measure at data point x_i is defined as

$$D_i = \sum_{j=1}^{n} exp\left(-\frac{\|x_i - x_j\|^2}{(r_a/2)^2}\right) \quad (4)$$

where r_a is a positive constant. Hence, a data point will have a high density value if it has many neighboring data points. The radius r_a defines a neighborhood for a data point, Data points outside the radius r_a contribute only slightly to the density measure.

After calculating the density measure of all data points, the data point with the highest density measure is selected as the first cluster center. Let x_{c1} be the point selected and D_{c1} is its corresponding density measure. The density measure Di, for each data point x_i are recalculated as follows:

$$D_i = D_i - D_{C1}\, exp\left(-\frac{\|x_i - x_{c1}\|^2}{\left(r_b/2\right)^2}\right) \quad (5)$$

Where, r_b is a positive constant. Therefore, data points close to the first cluster center x_{c1} will have significantly reduced density measure and are unlikely to be selected as the next cluster center. The constant r_b defines a neighborhood that has a measurable reduction in the density measure. The constant r_b is normally larger than r_a to prevent closely spaced cluster centers; generally r_b is equal to *1.5* r_a, as suggested in [25].

After the density measure for each data point is recalculated, the next cluster center x_{c2} is selected and the density measures for all data points are recalculated. This iterative process is repeated until a sufficient number of cluster centers are generated.

When applying subtractive clustering to a set of input-output data, each of the cluster centers represent a prototype that exhibits certain characteristics of the system to be modeled. These cluster centers would be reasonably used as the initial clustering centers for PSO algorithm.

B. PSO Algorithm

PSO was originally developed by Eberhart and Kennedy in 1995 based on the phenomenon of collective intelligence inspired by the social behavior of bird flocking or fish schooling [11]. In the PSO algorithm, the birds in a flock are symbolically represented by particles. These particles can be considered as simple agents "flying" through a problem space. A particle's location in the multi-dimensional problem space represents one solution for the problem. When a particle moves to a new location, a different problem solution is generated. The fitness function is evaluated for each particle in the swarm and is compared to the fitness of the best previous position for that particle p_{best} and to the fitness of the global best particle among all particles in the swarm g_{best}. After finding the two best values, the i^{th} particles evolve by updating their velocities and positions according to the following equations:

$$v_{id} = w * v_{id} + c_{1*} rand_1 * (p_{best} - x_{id}) + c_2 * rand_2 * (g_{best} - x_{id}) \quad (6)$$

$$x_{id} = x_{id} + v_{id} \quad (7)$$

where d denotes the dimension of the problem space; $rand_1$, $rand_2$ are random values in the range of (0, 1). The random values, $rand_1$ and $rand_2$, are used for the sake of completeness, that is, to make sure that particles explore wide search space before converging around the optimal solution. c_1 and c_2 are constants and are known as acceleration coefficients; The values of c_1 and c_2 control the weight balance of p_{best} and g_{best} in deciding the particle's next movement. w denotes the inertia weight factor; An improvement to original PSO is constituted by the fact that w is not kept constant during execution; rather, starting from a maximal value, it is linearly decremented as the number of iterations increases down to a minimal value [4], initially set to 0.9, decreasing to 0.4 according to:

$$w = (w - 0.4)\frac{(MAXITER - ITERATION)}{MAXITER} + 0.4 \quad (8)$$

MAXITER is the maximum number of iterations, and *ITERATION* represents the current number of iterations. The inertia weight factor w provides the necessary diversity to the swarm by changing the momentum of particles to avoid the stagnation of particles at the local optima. The empirical research conducted by Eberhart and Shi shows improvement of search efficiency through gradually decreasing the value of inertia weight factor from a high value during the search.

Equation 6 indicates that each particle records its current coordinate x_{id}, and its velocity v_{id} that indicates the speed of its movement along the dimensions in a problem space. For every generation, the particle's new location is computed by adding the particle's current velocity, v -vector, to its location, x -vector.

The best fitness values are updated at each generation, based on

$$P_i(t+1) = \begin{cases} P_i(t) & f(X_i(t+1)) \le f(X_i(t)) \\ X_i(t+1) & f(X_i(t+1)) > f(X_i(t)) \end{cases} \quad (9)$$

It is possible to view the clustering problem as an optimization problem that locates the optimal centroids of the clusters rather than finding an optimal partition [18, 20, 21, 22]. This view offers us a chance to apply PSO optimal algorithm on the clustering solution. The PSO clustering algorithm performs a globalized search in the entire solution space [4, 17]. Utilizing the PSO algorithm's optimal ability, if given enough time, the proposed hybrid Subtractive+(PSO) clustering algorithm can yield more compact clustering results compared to traditional PSO clustering algorithm. However, in order to cluster the large datasets, PSO requires much more iteration (generally more than 500 iterations) to converge to the optima than the hybrid Subtractive + (PSO) clustering algorithm does. Although the PSO algorithm is inherently parallel and can be implemented using parallel hardware, such as a computer cluster, the computation requirement for clustering extremely huge datasets is still high. In terms of execution time, hybrid Subtractive + (PSO) clustering algorithm is the most efficient for large dataset [1].

IV. HYBRID SUBTRACTIVE + (PSO) CLUSTERING ALGORITHM

In the hybrid Subtractive + (PSO) clustering algorithm, the multidimensional vector space is modeled as a problem space. Each vector can be represented as a dot in the problem space. The whole dataset can be represented as a multiple dimension space with a large number of dots in the space. The hybrid Subtractive + (PSO) clustering algorithm includes two modules, the Subtractive clustering module and PSO module. At the initial stage, the Subtractive clustering module is executed to search for the clusters' centroid locations and the suggested number of clusters. This information is transferred to the PSO module for refining and generating the final optimal clustering solution.

The Subtractive clustering module: The Subtractive clustering module predicts the optimal number of clusters and finds the optimal initial cluster centroids for the next phase.

The PSO clustering module: In the PSO clustering algorithm, the whole dataset can be represented as a multiple dimension space with a large number of dots in space. One particle in the swarm represents one possible solution for clustering the dataset. Each particle maintains a matrix $X_i = (C_1, C_2, \ldots, C_i, .., C_k)$, where C_i represents the i^{th} cluster centroid vector and k represent the total number of clusters. According to its own experience and those of its neighbors, the particle adjusts the centroid vector position in the vector space at each generation. The average distance of data objects to the cluster centroid is used as the fitness value to evaluate the solution represented by each particle. The fitness value is measured by the equation below:

$$f = \frac{\sum_{i=1}^{N_c} \left\{ \frac{\sum_{j=1}^{p_i} d(o_i, m_{ij})}{p_i} \right\}}{N_c} \qquad (10)$$

where m_{ij} denotes the j^{th} data vector, which belongs to cluster i ; o_i is the centroid vector of i^{th} cluster; $d(o_i,m_{ij})$ is the distance between data point m_{ij} and the cluster centroid o_i ; p_i stands for the data number, which belongs to $cluster_c$; N_c stands for the cluster number.

In the hybrid Subtractive + (PSO) clustering algorithm, the Subtractive algorithm is used at the initial stage to help discovering the vicinity of the optimal solution by suggesting good initial cluster centers and the number of clusters. The result from Subtractive algorithm is used as the initial seed of the PSO algorithm, which is applied for refining and generating the final result using the global search capability of PSO. The flow chart of the hybrid Subtractive + (PSO) is depicted graphically in Figure 1.

V. EXPERIMENTAL STUDIES

The main purpose in this paper is to compare the quality of the PSO and hybrid Subtractive + (PSO) clustering algorithm, where the quality of the clustering is measured according to the intra–cluster distances, i.e. the distance between the data vectors and the cluster centroid within a cluster, where the objective is to minimize the intra-cluster distances.

Clustering Problem: We used three different data collections to compare the performance of the PSO and hybrid Subtractive + (PSO) clustering algorithms. These datasets are downloaded from Machine Learning Repository site [23]. A description of the test datasets is given in Table 1. In order to reduce the impact of the length variations of different data, each data vector is normalized so that it is of unit length.

Table 1: Summary of datasets

	Number of Instances	**Number of Attributes**	**Number of classes**
Iris	150	4	3
Wine	178	13	3
Yeast	1484	8	10

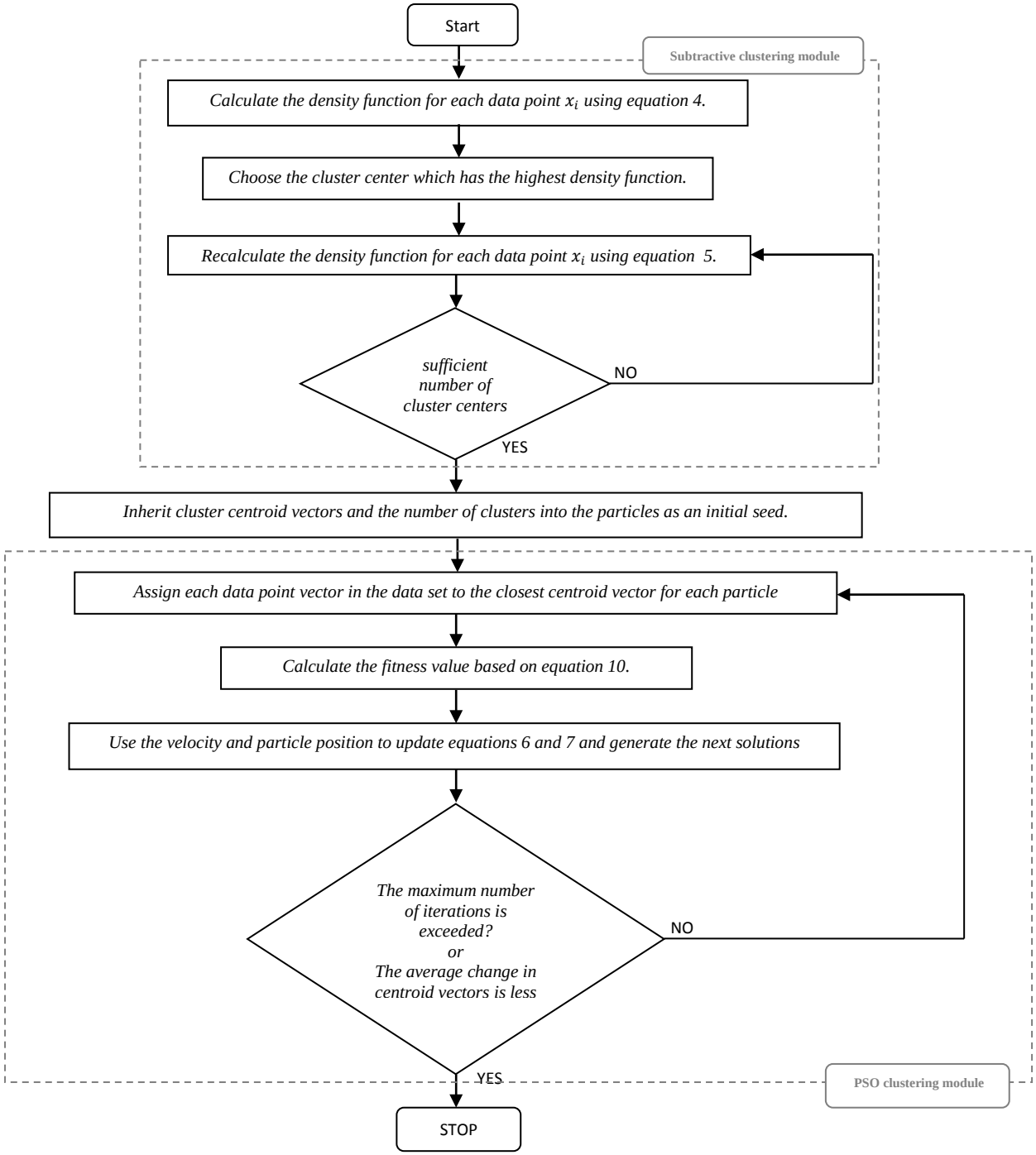

Figure 1: The flowchart of hybrid Subtractive + (PSO)

Experimental Setting: In this section we present the experimental results of the hybrid Subtractive + (PSO) clustering algorithm. For the sake of comparison, we also include the results of the Subtractive and PSO clustering algorithms. In our case the Euclidian distance measure is used as the similarity metrics in each algorithm. The performance of the clustering algorithm can be improved by seeding the initial swarm with the result of the subtractive algorithm. $c_1 = c_2 = 1.49$ and w inertia weight is according to equation (8). These values are chosen respectively based on the results reported in [17]. We choose number of particles as a function of number of classes. In Iris plants database we chose 15 particles, 15 particles in Wine database and 50 particles in the Yeast database. All these values were chosen to ensure good convergence.

Results and Discussion: The fitness equation (10) is used not only in the PSO algorithm for the fitness value calculation, but also in the evaluation of the cluster quality. It indicates the value of the average distance between a data point and the cluster centroid to which they belong. The smaller the value, the more compact the clustering solution is. *Table 2* demonstrates the experimental results by using the Subtractive, PSO and Subtractive + (PSO) clustering algorithm respectively. For an easy comparison, the PSO and hybrid Subtractive + (PSO) clustering

algorithm runs 200 iterations in each experiment. For all the result reported, averages over more than ten simulations are given in Table 2. To illustrate the convergence behavior of different clustering algorithms, the clustering fitness values at each iteration are recorded when these two algorithms are applied on datasets separately. As shown in Table 2, the Subtractive + (PSO) clustering approach generates the clustering result that has the lower fitness value for all three datasets using the Euclidian similarity metric, The results from the Subtractive+(PSO) approach have improvements compared to the results of the PSO approach.

Table 2: Performance comparison Subtractive, PSO, Subtractive+(PSO)

	Fitness value		
	Subtractive	**PSO**	**Subtractive + PSO**
Iris	6.12	6.891	3.861
Wine	2.28	2.13	1.64
Yeast	1.30	1.289	0.192

Figure 2 shows the suggested cluster centers done by the Subtractive clustering algorithm, the cluster centers appear in black as shown figure. These centers will be used as the initial seed of the PSO algorithm. Figures 3, 4, 5 illustrate the convergence behaviors of the two algorithms on the three datasets using the Euclidian distance as a similarity metric.

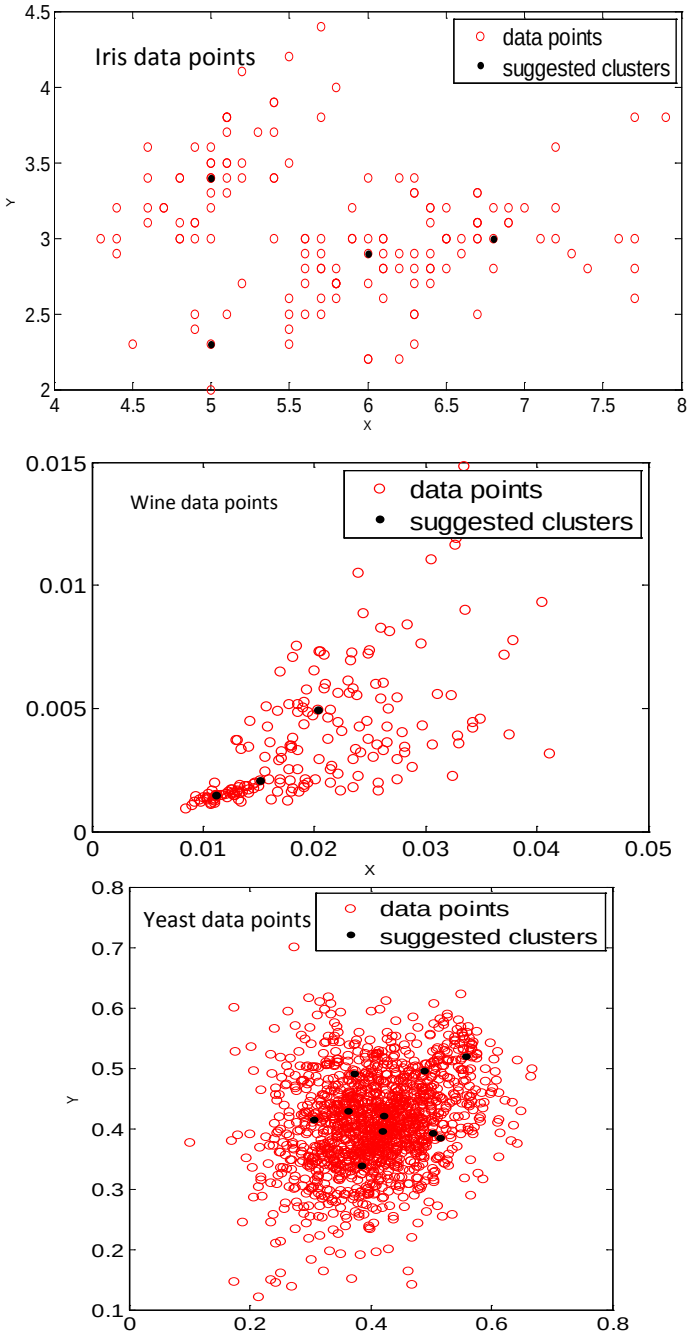

Figure 2: Suggested cluster centers by the Subtractive clustering algorithm

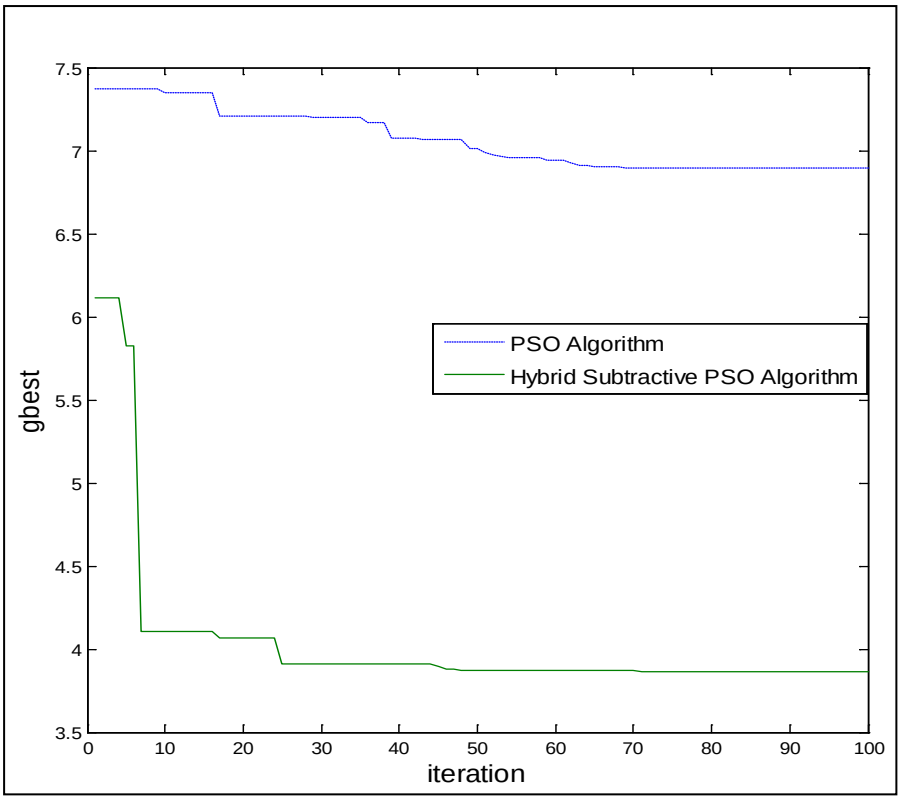

Figure 3: Algorithm convergence for Iris database

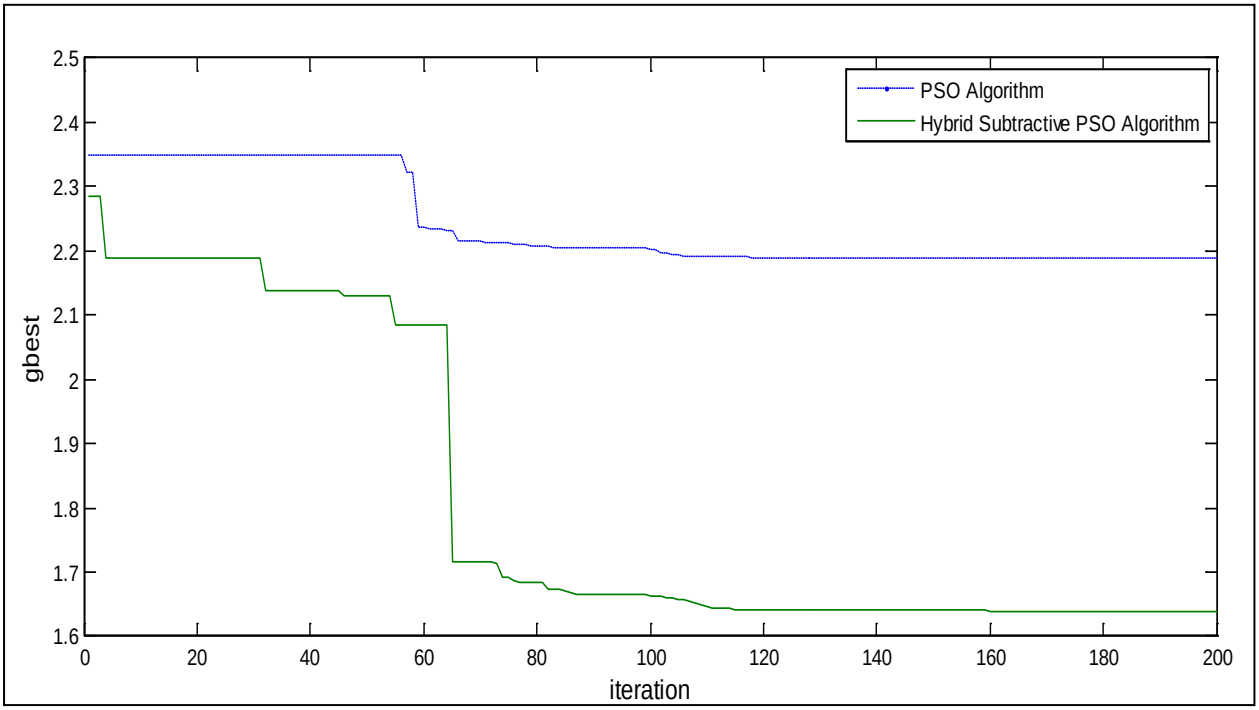

Figure 4: Algorithm convergence for Wine database

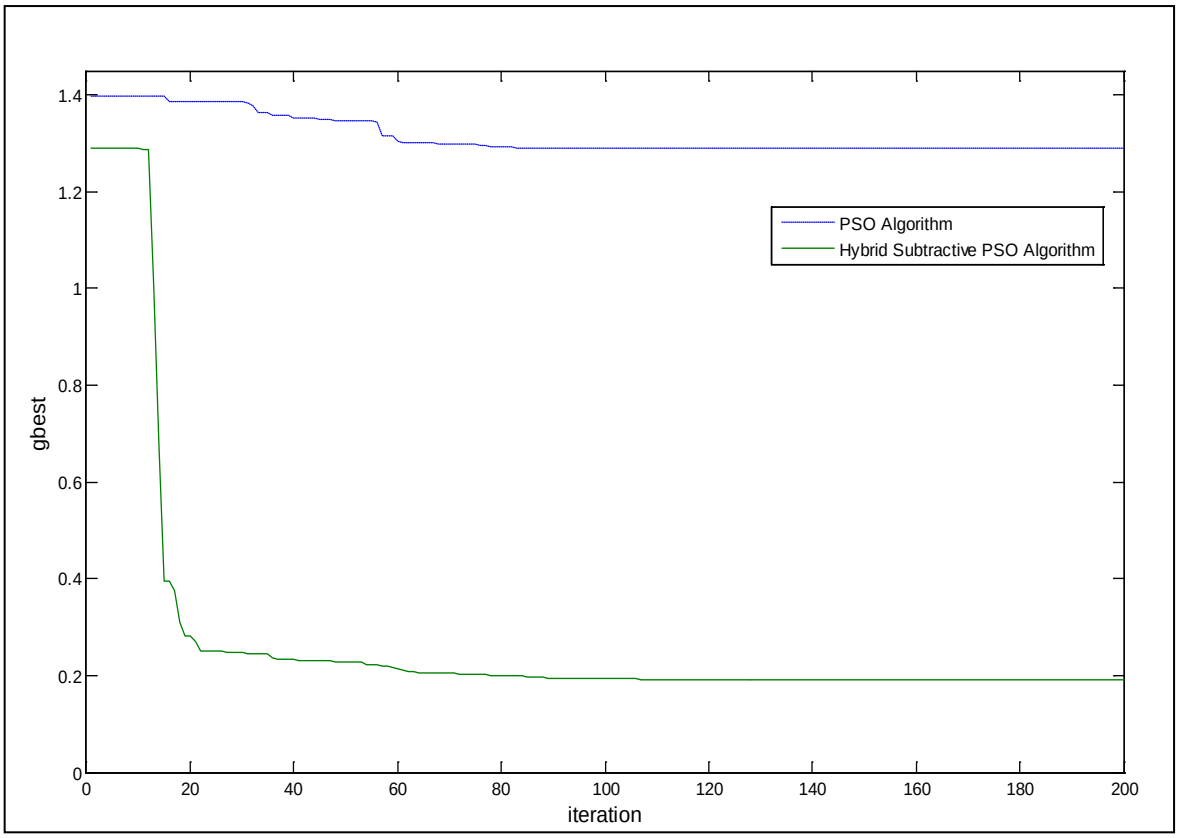

Figure 5: Algorithm convergence for Yeast database

In all the previous figures representing Hybrid Subtractive PSO Algorithm, g_{best} is the fitness of the global best particle among all particles in the swarm particle that was inherited with cluster centroid vectors from Subtractive algorithm. Now, we notice that the Subtractive + PSO algorithm has a good start and it converges quickly with lower fitness function. As shown in Figure 3, the fitness value of the Subtractive + PSO algorithm starts with 6.1 and it reduced sharply from 6.1 to 3.8 within 25 iterations and fixed at 3.68. The PSO algorithm starts with 7.4, the reduction of the fitness value in PSO is not as sharp as in Subtractive + PSO and becomes smoothly after 55 iterations. The same happened in the Figures 4 and 5. The Subtractive + PSO algorithm shows good improvement for large dataset as shown in Figure 5. This indicates that upon termination the Subtractive + PSO yield minimal fitness values. Therefore, the proposed algorithm is an efficient and effective solution to the data clustering problem.

VI. CONCLUSION

This paper investigated the application of the Subtractive + PSO algorithm, which is a hybrid of PSO and Subtractive algorithms to cluster data vectors. Subtractive clustering module is executed to search for the cluster's centroid locations and the suggested number of clusters. This information is transferred to the PSO module for refining and generating the final optimal clustering solution. In the general PSO algorithm, PSO can conduct a globalized searching for the optimal clustering, but it requires more iteration numbers. The subtractive clustering helps the PSO to start with good initial cluster centroid to converge faster with small fitness function which means a more compact result. The algorithm includes two modules, the Subtractive module and the PSO module. The Subtractive module is executed at the initial stage to discover good initial cluster centroids. The result from the Subtractive module is used as the initial seed of the PSO module to discover the optimal solution by a global search and at the same time to avoid consuming high computation. The PSO algorithm will be applied for refining and generating the final result. Experimental results illustrate that using this hybrid Subtractive + PSO algorithm can generate better clustering results compared to using ordinary PSO.

VII. REFERENCES

[1] Khaled S. Al-Sultana, M. Maroof Khan, "Computational experience on four algorithms for the hard clustering problem". Pattern Recognition Letter, Vol.17, No.3, pp.295–308, 1996. doi: 10.1016/0167-8655(95)00122-0

[2] Michael R. Anderberg , "Cluster Analysis for Applications". Academic Press Inc., New York, 1973.

[3] Pavel Berkhin, "Survey of clustering data mining techniques". Accrue Software Research Paper, pp.25-71, 2002. doi: 10.1007/3-540-28349-8_2

[4] A. Carlisle, G. Dozier, "An Off-The- Shelf PSO". In Proceedings of the Particle Swarm Optimization Workshop, 2001, PP: 1-6.

[5] Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski, "Data Mining – Methods for Knowledge Discovery". Kluwer Academic Publishers, 1998. doi: 10.1007/978-1-4615-5589-6

[6] X. Cui, P. Palathingal, T.E. Potok, "Document Clustering using Particle Swarm Optimization". IEEE Swarm Intelligence Symposium 2005, Pasadena, California, pp. 185 - 191. doi: 10.1109/SIS.2005.1501621

[7] Eberhart, R.C., Shi, Y. "Comparing Inertia Weights and Constriction Factors in Particle Swarm Optimization". Congress on Evolutionary Computing, vol. 1, 2000, pp: 84-88. doi: 10.1109/CEC.2000.870279

[8] Everitt, B. "Cluster Analysis". 2nd Edition, Halsted Press, New York, 1980.

[9] A. K. Jain , M. N. Murty , P. J. Flynn, "Data Clustering: A Review". ACM ComputingSurvey, Vol. 31, No. 3, pp: 264-323, 1999. doi: 10.1145/331499.331504

[10] J. A. Hartigan, "Clustering Algorithms". John Wiley and Sons, Inc., New York, 1975.

[11] Eberhart RC, Shi Y, Kennedy J, "Swarm Intelligence". Morgan Kaufmann, New York, 2001.

[12] Mahamed G. Omran, Ayed Salman, Andries P. Engelbrecht, "Image classification using particle swarm optimization". Proceedings of the 4th Asia-Pacific Conference on Simulated Evolution and Learning 2002, Singapore, pp: 370-374. doi: 10.1142/9789812561794_0019

[13] S. L. Chiu, "Fuzzy model identification based on cluster estimation". Journal of Intelligent and Fuzzy Systems, Vol. 2, No. 3, 1994.

[14] Salton G. and Buckley C., "Term-weighting approaches in automatic text retrieval". Information Processing and Management, Vol. 24, No. 5, pp: 513-523, 1988. doi: 10.1016/0306-4573(88)90021-0

[15] Song Liangtu, Zhang Xiaoming, "Web Text Feature Extraction with Particle Swarm Optimization". IJCSNS International Journal of Computer Science and Network Security, Vol. 7, No. 6, 2007.

[16] Selim, Shokri Z., "K-means type algorithms: A generalized convergence theorem and characterization of local optimality". Pattern Analysis and Machine Intelligence, IEEE Transactions Vol. 6, No.1, pp:81–87, 1984. doi: 10.1109/TPAMI.1984.4767478

[17] Yuhui Shi, Russell C. Eberhart, "Parameter Selection in Particle Swarm Optimization". The 7th Annual Conference on Evolutionary Programming, San Diego, pp. pp 591-600, 1998. doi: 10.1007/BFb0040810

[18] Michael Steinbach, George Karypis, Vipin Kumar, "A Comparison of Document Clustering Techniques". TextMining Workshop, KDD, 2000.

[19] Razan Alwee, Siti Mariyam, Firdaus Aziz, K.H.Chey, Haza Nuzly, "The Impact of Social Network Structure in Particle Swarm Optimization for Classification Problems". International Journal of Soft Computing , Vol. 4, No. 4, 2009, pp:151-156.

[20] Van D. M., Engelbrecht. A.P., "Data clustering using particle swarm optimization". Proceedings of IEEE Congress on Evolutionary Computation 2003, Canbella, Australia. pp: 215-220. doi: 10.1109/CEC.2003.1299577

[21] Sherin M. Youssef, Mohamed Rizk, Mohamed El-Sherif, "Dynamically Adaptive Data Clustering Using Intelligent Swarm-like Agents". International Journal of Mathematics and Computer in simulation, Vol. 1, No.2, 2007.

[22] Rehab F. Abdel-Kader, "Genetically Improved PSO Algorithm for Efficient Data Clustering". Proceeding Second International Conference on Machine Learning and Computing 2010, pp.71-75. doi: 10.1109/ICMLC.2010.19

[23] UCI Repository of Machine Learning Databases. http://www.ics.uci .edu/~mlearn/MLRepository.html .

[24] K. Premalatha, A.M. Natarajan, "Discrete PSO with GA Operators for Document Clustering". International Journal of Recent Trends in Engineering, Vol. 1, No. 1, 2009.

[25] JunYing Chen, Zheng Qin, Ji Jia, "A Weighted Mean Subtractive Clustering Algorithm". Information Technology Journal, No. 7, pp.356-360, 2008. doi: 10.3923/itj.2008.356.360

[26] Neveen I. Ghali, Nahed El-dessouki, Mervat A. N, Lamiaa Bakraw, "Exponential Particle Swarm Optimization Approach for Improving Data Clustering". International Journal of Electrical & Electronics Engineering, Vol. 3, Issue 4, May 2009.

[27] Vijay Kalivarapu, Jung-Leng Foo, Eliot Winer, "Improving solution characteristics of particle swarm optimization using digital pheromones". Structural and Multidisciplinary Optimization - STRUCT MULTIDISCIP OPTIM, Vol. 37, No. 4, pp: 415-427, 2009. doi: 10.1007/s00158-008-0240-9

[28] H. Izakian, A. Abraham, and V. Snásel, "Fuzzy Clustering using Hybrid Fuzzy c-means and Fuzzy Particle Swarm Optimization", World Congress on Nature & Biologically Inspired Computing, NaBIC 2009. In Proc. NaBIC, pp.1690-1694, 2009. doi: 10.1109/NABIC.2009.5393618

How to cite

Mariam El-Tarabily, Rehab Abdel-Kader, Mahmoud Marie, Gamal Abdel-Azeem, " A PSO-Based Subtractive Data Clustering Algorithm ". *International Journal of Research in Computer Science*, 3 (2): pp. 1-9, March 2013. doi: 10.7815/ijorcs. 32.2013.060