International Journal of Systems Signal Control and Engineering Application 10 (1-6): 41-47, 2017
ISSN: 1997-5422

Improved Particle Swarm Optimization Algorithm in K-Means

[1]Y. Farhang, [2]A. Afroozeh and [3]K. Jahanbin
[1]Faculty of Computer, Islamic Azad University, Khoy Branch, Khoy, Iran
[2]Larestan Higher Education Complex, Larestan, Iran
[3]Department of Computer, Kerman Branch, Islamic Azad University, Kerman, Iran

Abstract: In recent years, combinational optimization issues are introduced as critical problems in clustering algorithms to partition data in a way that optimizes the performance of clustering. K-means algorithm is one of the famous and more popular clustering algorithms which can be simply implemented and it can easily solve the optimization issue with less extra information. In this regard, researchers have worked to improve the problem computationally, creating efficient solutions that lead to better data analysis through the K-means Clustering algorithm. Finally, the Partial Swarm Optimization (GAPSO) and Partial Swarm Optimization-Genetic Algorithm (PSOGA) through the K-means algorithm were proposed.

Key words: Improved particle, swarm optimization, algorithm, PSOGA, K-means, clustering

INTRODUCTION

One of the important and constantly developing issues in the world of science is Computer Science (CS) which is the practical and scientific approach used for computation and its related applications. CS studies systematizes the mechanization, feasibility, expression and structure of methodical algorithms that underlie the acquisition, processing, representation, storage, access to and communication of information. In recent years, to solve the data clustering problem, several new approaches have been introduced, inspired from biological sciences including Genetic algorithm, particle swarm optimization algorithm and so on Amiri *et al*. (2011, 2015) and Afroozeh *et al*. (2011, 2013). Also, existing Hybrid algorithms with K-means clustering suffer from different drawbacks such as lack of providing optimum solution for all problems, getting stuck in local optima, tuning many parameters, slow convergence rate, high number of error and high intra cluster distance. Also, existing Meta-Heuristic algorithms with K-means clustering have low accuracy rate of the clustering and low the number of correct answers, they have good performance only in one of the search spaces (Afroozeh *et al*., 2014a, 2015; Amiri and Afroozeh, 2014a-c; Jalil *et al*., 2012). However, the algorithms are robust and have the ability of adapting with changing environment. Due to the shortcomings of the K-means Clustering algorithm, it can be optimized when used in the form of Hybrid algorithms. Two popular algorithms that are mostly used in Hybrid algorithms due to their high-performance are Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) algorithm. Given that these two algorithms do not have a label and additional guides, they can be used to improve the performance of K-means algorithms. In the following sections, GA, PSO and hybrid of them with the K-means clustering algorithm is discussed (Bahadoran *et al*., 2012; Afroozeh *et al*., 2011, 2013, 2014b).

MATERIALS AND METHODS

Particle Swarm Optimization (PSO) method: Other hybrid method used in this study is Particle Swarm Optimization (PSO) method that is explained in this part. Furthermore, this study contains three major parts: definition of PSO, description of PSO and PSO for Clustering algorithm.

Definition of Particle Swarm Optimization (PSO): Particle Swarm Optimization (PSO) is a global optimization algorithm that is employed to address the problems wherein a best solution can be denoted as a surface or point within an n-dimensional space in which hypotheses are plotted and seeded with aninitial velocity together with a communication channel between particles (Gomez *et al*., 2010; Sadeghierad *et al*., 2010; Shi and Eberhart, 1998; Kennedy, 1997; Urade and Patel, 2012).

Afterwards, the particles move throughout the solution space and subsequent to each time step, they are

Corresponding Author: Y. Farhang, Faculty of Computer, Islamic Azad University, Khoy Branch, Khoy, Iran

Afterwards, the particles move throughout the solution space and subsequent to each time step, they are assessed based on some fitness criterion (Robinson and Rahmat-Samii, 2004). During time, particles are speeded up toward the particles that exist in their communication grouping that have better fitness values. The major benefit of this approach compared to other strategies of global minimization, e.g., simulated annealing is that large number of members that form the particle swarm make this technique notably flexible to the local minima problem (Sadeghierad *et al*., 2010; Shi and Eberhart, 1998; Kennedy, 1997; Urade and Patel, 2012; Shakerian *et al*., 2011).

PSO which is a search algorithm based on population, gets started with a population of random solutions that are known as particles (Hu *et al*., 2004). In PSO, each solution is similar to a 'bird' within the searchspace which is known as 'particle'. All particles are associated with fitness values which are assessed by fitness function to be optimized as well as velocities that direct the flying of particles. The particles fly throughout the problem space and follow those particles that have the best solutions. The initialization of PSO is with a group of random particles then, it searches for optima through updating each generation (Premalatha and Natarajan, 2010).

Omran *et al*. (2002) introduced a straightforward implementation of PSO algorithm for clustering. Their algorithm employed a fixed number of clusters and applied PSO to searching for the best centroids of the clusters.

Merwe and Engelbrecht (2003) introduced two new approaches by means of PSO to be applied to cluster data. They showed how PSO could be employed for finding the centroids of a user-specified number of clusters. Then, the algorithm was extended to employ K-means algorithm for seeding the initial swarm. The second algorithm applied PSO to refining the clusters that were created by K-means. The performance of new PSO algorithms are assessed on six datasets and the result is compared to the K-means clustering performance (Abul Hasan and Ramakrishnan, 2011).

The second part of the study deals with the design and development of research composed of three phases. In the first phase, a Genetic Algorithm (GA) is used for enhancement of the K-means clustering algorithm. Therefore, in this phase, e GA-K-means algorithm is improved in order to decrease the error in K-means algorithm this is called Improved Genetic algorithm in K-means algorithm (I-GA-K-means algorithm). In the second phase, Particle Swarm Optimization (PSO) algorithm is applied to enhance the K-means Clustering algorithm. Thus, in this phase, the PSO-K-means algorithm is improved in a way to decrease the intra-cluster distance in K-means algorithm this is called Improved Particle Swarm Optimization in the K-means algorithm (I-PSO-K-means algorithm). In the third phase, two Meta-Heuristic algorithms of the two previous Hybrid algorithms (I-GA-K-means and I-PSO-K-means) are proposed for the enhancement of K-means Clustering algorithm. Consequently, the two algorithms are proposed to increase the accuracy rate in K-means algorithm they are called the improved Genetic algorithm the Improved Particle Swarm Optimization in K-means algorithm (GAPSO-K-means algorithm) and the Improved Particle Swarm Optimization the Improved Genetic algorithm in K-means algorithm (PSOGA-K-means algorithm).

Secondly, the method of second proposed algorithm (Improved Particle Swarm Optimization-K-means) is fully described and then the results of this proposed algorithm will be expressed. The results of the Improved Particle Swarm Optimization-K-means (I-PSO-K-means) algorithm are obtained and then compared to the results obtained from previous algorithms and PSO-K-means algorithm. In this phase, the K-means algorithm (Meila, 2006) and the PSO-K-means clustering algorithm (Tsai and Kao, 2011) are used.

Modeling of I-PSO-K-means algorithm: This study improves the particle swarm optimization algorithm in K-means algorithm. Additionally, this study addresses the second objective of the study. One of the shortcomings of the PSO-K-means clustering algorithm is the high intra-cluster distance in the clustering of datasets which can be low. To this end, the Improved Particle Swarm Optimization-K-means algorithm (I-PSO-K-means) is proposed. In the following, the design of I-PSO-K-means algorithm is described. The proposed algorithm in this study comprises eight important steps: initialization, compare for obtaining Pbeast, compare for obtaining Gbeast, calculating the function, checking the Max-domain, checking the min-domain and checking the repeat and running K-means (Fig. 1).

Here, the implementation ofthe I-PSO-K-means clustering algorithm (Improved Particle Swarm Optimization-K-means algorithm) is elaborated. As mentioned in the previous study, the algorithm proposed here is a hybrid of the PSO algorithm and the K-means clustering algorithm. In the following, the implementation of the I-PSO-K-means clustering algorithm is described. The proposed algorithm in this study has fourteen main steps. These steps are shown in Algorithm 1.

Algorithm 1 (Improved particle swarm optimization-K-means):

1. Start
2. Selection the Initial Population.
   a. Appling Dataset
   b. Finding Domain of All Attributes in Dataset (h_1, h_2, …)
      i. Finding Minimum Domain ($h_1$min, $h_2$min, …)
      ii. Finding Minimum Domain ($h_1$max, $h_2$max, …)
      iii. Finding Domain ($h_1 = h_1$max-$h_1$min, $h_2 = h_1$max-$h_2$min, …)
      iv. For f = 1 to 50
         1. Finding 50 Cluster Centers by Randomly in Domain (Initial Population)
         2. Running K-means Algorithm on Initial Population
            a. For n = 1 to 50
               i. Mining of Features nth Cluster Center from Dataset (m_1, m_2, …)
               ii. For K = 1 to N*/N is No. members in Dataset. /*
                  1. Mining of Features Kth Row from Dataset (a_1, a_2, …)
                  2. Calculate Distance (m_1, m_2, …) and (a_1, a_2, …) by Euclidean
                  3. Finding of Minimum Distance between Cluster Centers
                  4. Placement in Cluster that it has Minimum Distance
                  5. Calculate total of distance (S_1, S_2, …)
               iii. end For
               iv. Calculate S ($S = S_1+S_2+$ …) */ S is Intra-cluster Distance. /*
               v. Placement Cluster Centers and S in the Matrix.(Matrix Name is P)
            b. end For
         3. Evaluation
            a. Sorting Rows of Matrix P Based on S (Descending P is P_1)
         4. Selection
            a. Selection one members from P_1 for Next Section. (Lowest S)
         5. Placement of Previously member in P_0 (P_0 (f, 1:Co))
      v. end For
3. Evaluation
   a. Placement P_0 in P (P is 50 members with S minimum)
   b. Sorting Rows of Matrix P Based on S (Descending P is P_1)
4. Selection
   a. Selection first members from P_1 for Gbest. (Lowest S)
5. Initial Values
   a. Gbest = P_1 (1,1:Co);
   b. W = 0.7299;
   c. C_1 = 1.4963;
   d. C_2 = 1.4963;
   e. R_1 = Random (0 to 1)
   f. R_2 = Random (0 to 1)
6. Specified Pbest
   a. Pbest = P(1,1:Co);
7. Specified Xt
   a. Xt = P(1,1:Co);
8. Vt = 0;
9. PSO Operator
   a. For f = 1 to 50
      i. Vtt = (W*Vt)+(C1*R1)(Pbest-Xt)+(C2*R2)(Gbest-Xt)
      ii. Xtt = Xt+Vtt
      iii. Checking Xtt
         1. If (Xtt>Max_ Max_ Domain)
            a. Xtt = Max_ Domain
         2. If (Min_ Domain>Xtt)
            a. Xtt = Min_ Domain
         3. Checking for All Columns
      iv. Xt = Xtt;
      v. Vt = Vtt
      vi. S = 0; (S = 0, S_1 = 0, S_2 = 0, …)
      vii. Placement Xtt on P_2
      viii. Running K-means Algorithm on Xtt
         1. For K = 1 to N*/N is number of members in Dataset. /*
            a. Mining of Features Kth Row from Dataset (a_1, a_2, …)
            b. Calculate Distance (m_1, m_2, …) and (a_1, a_2, …) by Euclidean
            c. Finding of Minimum Distance between Cluster Centers
            d. Placement in Cluster that it has Minimum Distance
            e. Calculate total of distance (S_1, S_2, …)
         2. end For
         3. Calculate S ($S = S_{1+} S_{2+ \ldots}$) */ S is Intra-cluster Distance. /*
         4. Placement Cluster Centers and S in the Matrix.(Matrix Name is P_3)
         5. Evaluation
            a. Sorting Rows of Matrix P_3 Based on S (Descending P_3 is P_4)
         6. Selection
            a. Selection 50 members from P_4 for Next Section. (Lowest S)

ix. Selection
1. Placement Cluster Centers with intra-cluster centers P_1 on P_5 (50)
2. Placement Cluster Centers with intra-cluster centers P_3 on P_5 (50)

x. Evaluation
1. Sorting Rows of Matrix P_5 Based on S (Descending P_5 is P_6)

xi. Specified Pbest
1. Pbest = P_4 (1,1:Co);

xii. Specified Gbest
1. Gbest = P_6 (1,1:Co);

b. end For

10. Iteration of Steps 6 to 9 (50 times).
11. Selection
    a. Selection first members from P_6 for Final Answer.
12. Running K-means Algorithm on Final Answer
    a. For K = 1 to N*/N is number of members in Dataset. /*
        i. Mining of Features Kth Row from Dataset (a_1, a_2, …)
        ii. Calculate Distance (m_1, m_2, …) and (a_1, a_2, …)by Euclidean
        iii. Finding of Minimum Distance between Cluster Centers
        iv. Placement in Cluster that it has Minimum Distance
        v. Calculate total of distance (S_1, S_2, …)
    b. end For
    c. Calculate S ($S = S_1+S_2+ \ldots$) */S is Intra-cluster Distance. /*
13. Drawing of Chart
14. Stop

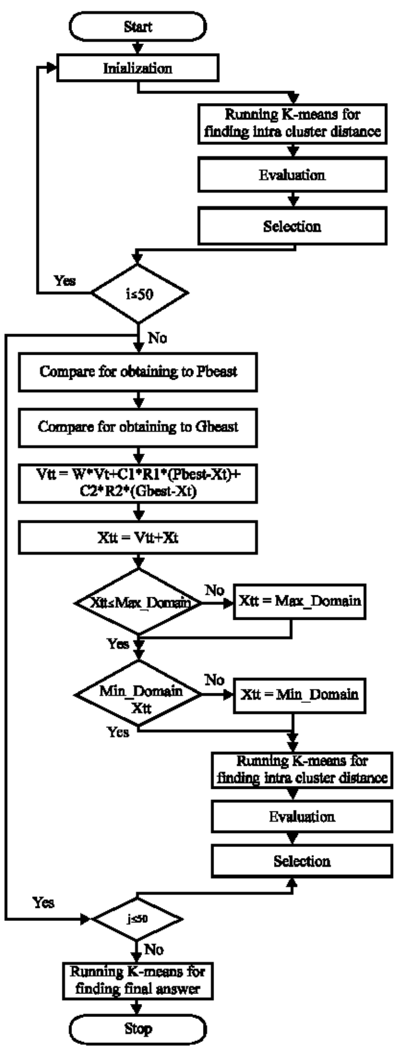

Fig. 1: The flowchart of I-PSO-K-means algorithm

In the I-PSO-K-means Clustering algorithm, there is innovation in different parts of the algorithm. This algorithm is a hybrid of the K-means Clustering algorithm and the Particle Swarm Optimization algorithm which reduces the intra-cluster distance in the K-means Clustering algorithm. In the next study, the first proposed algorithm is investigated using different datasets and the results are compared with those of other algorithms.

RESULTS AND DISCUSSION

This study explains the results obtained from the I-PSO-K-means algorithm (Improved Particle Swarm Optimization-K-means). This algorithm is fully described in the previous study. In this study, the proposed algorithms are analyzed using six standard data sets explained in Chapter 3 (i.e., Balance, Blood, Breast, Iris, Pima and Wine). This study is organized into two main parts: analysis and discussion of the I-PSO-K-means algorithm.

In this part, the results of I-PSO-K-means clustering algorithm are discussed. Since, the I-PSO-K-means algorithm is related to the second phase of the study in this phase, the comparison factor is intra-cluster distance. Therefore, in this study two important areas, namely the intra-cluster distance that is the average of intra-cluster distance and the standard deviation of intra-cluster distance are analyzed. In Fig. 2, the average of intra-cluster distance is shown for 20 times of running of the five algorithms.

In Fig. 3, it can be seen that the average of intra-cluster distance in the proposed algorithm

(Improved Particle Swarm Optimization-K-means) is better than the previous algorithms. Therefore, the performance of the proposed algorithm in this phase can be better than previous algorithms.

Finally, the Improved Particle Swarm Optimization-K-means algorithm (I-PSO-K-means) was examined to solve clustering problems. The I-PSO-K-means algorithm was improved to the PSO-K-means algorithm; the I-PSO-K-means algorithm reduced the intra-cluster distance that it is related to the second objective. The second proposed method was applied to six UCI standard data sets and the result was contrasted with K-means, PSO-K-means and previous algorithms. The proposed algorithms in this chapter improved the K-means algorithm and solved some clustering problems. In this study, two meta-heuristic optimization algorithms will be proposed for development of K-means clustering algorithm.

The main contribution of this thesis is proposing Hybrid algorithms and new Meta-Heuristic algorithms to enhance learning algorithms for K-means clustering algorithm. In the next studies, the proposed Meta-Heuristic methods, i.e., improved Genetic algorithm improved particle swarm optimization in K-means and improved particle swarm optimization improved Genetic algorithm in K-means are introduced. These algorithms are an improved scheme of K-means clustering based on GA and PSO. The GA and PSO laws which are used to design the proposed methods are described in the following section and the proposed algorithms and their results for function optimization are provided in the next sections of this chapter.

Figure 3, it can be observed that the average number of errors in the proposed algorithm (Improved Particle Swarm Optimization and Improved Genetic Algorithm-K-means) is better than the previous algorithm. Thus, the PSOGA-K-means algorithm in this phase will have better performance than previous algorithms.

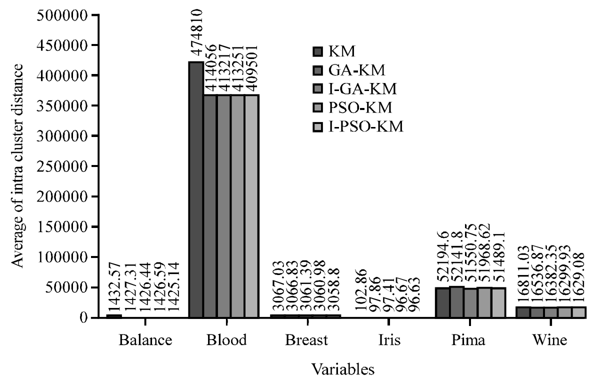

Fig. 2: The average of intra-cluster distance in I-PSO-KM

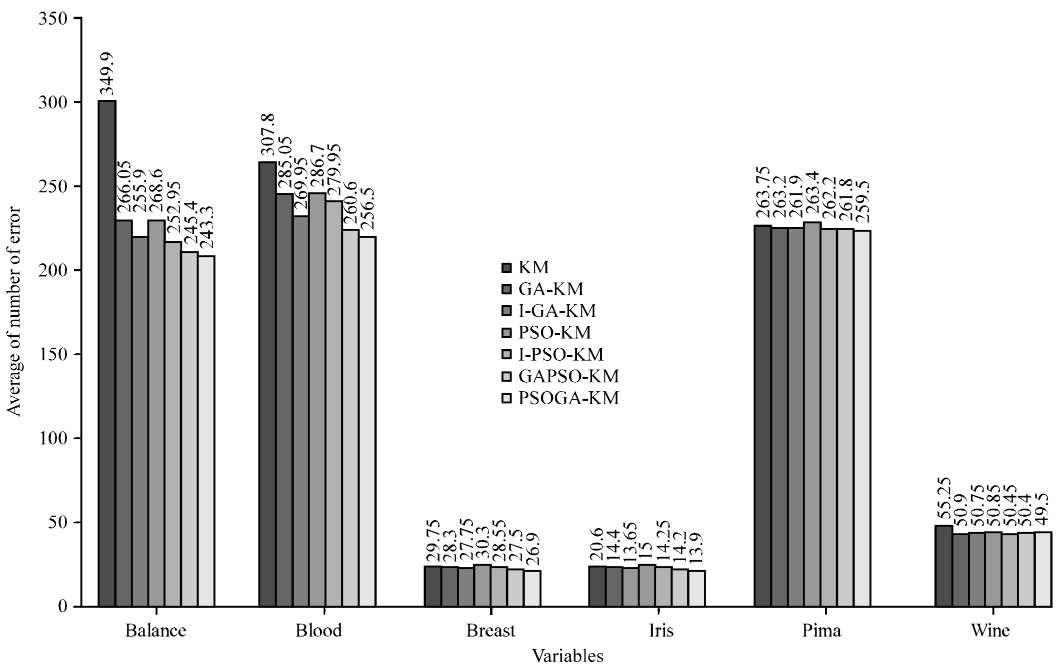

Fig. 3: The average of number of error in all proposed algorithms

CONCLUSION

This study proposed new Hybrid algorithms and new Meta-Heuristic algorithms and an improved scheme of K-means for solving clustering problems. The algorithms were named Improved Genetic Algorithm-K-means (I-GA-K-means), Improved Particle Swarm Optimization (I-PSO-K-means), improved Genetic Algorithm-improvedParticle Swarm Optimization-K-means (GAPSO-K-means) and improved Particle Swarm Optimization-improved Genetic Algorithm-K-means (PSOGA-K-means). The aim of these algorithms was to accelerate the learning, increase accuracy, decrease intra-cluster distance and decrease error in solving clustering problems. The ability of these algorithms was studied using UCI standard data sets in clustering problems. The results showed that the performance of proposed methods is better than previous methods in the clustering.

REFERENCES

Abul Hasan, M.J. and S. Ramakrishnan, 2011. A survey: Hybrid evolutionary algorithms for cluster analysis. Artif. Intell. Rev., 36: 179-204.

Afroozeh, A., I.S. Amiri, K. Chaudhary, J. Ali and P.P. Yupapin, 2014a. Analysis of optical ring resonator. J. Opt. Res., 16: 135-147.

Afroozeh, A., S.E. Pourmand and A. Zeinalinezhad, 2014b. Simulation and calculation of the parameters of PCF under terahertz wave band. Adv. Environ. Biol., 28: 263-269.

Afroozeh, A., I.S. Amiri, S.E. Pourmand, A. Zeinalinezhad and S.E. Alavi *et al.*, 2015. Comparison of control light using kramers-kronig method by three waveguides. J. Comput. Theor. Nanosci., 12: 1864-1868.

Afroozeh, A., K. Innate, J. Ali and P.P. Yupapin, 2013. THz frequency generation using Gaussian pulse for medical applications. Optik. Int. J. Light Elect. Optics, 124: 416-419.

Afroozeh, A., R. Jomtarak, J. Ali and P.P. Yupapin, 2011. High-terahertz-frequency carrier generation by optical pulse for radio-over-fiber applications. Opt. Eng., 50: 125005-125011.

Amiri, I.S. and A. Afroozeh, 2014a. Integrated Ring Resonator Systems. In: Ring Resonator Systems to Perform Optical Communication Enhancement Using Soliton, Amiri, I.S. and A. Afroozeh (Eds.). Springer, Berlin, Germany, ISBN:978-981-287-196-1, pp: 37-47.

Amiri, I.S. and A. Afroozeh, 2014b. Ring Resonator Systems to Perform Optical Communication Enhancement using Soliton. Springer, Berlin, Germany, ISBN:978-981-287-196-1, Pages: 69.

Amiri, I.S. and A. Afroozeh, 2014c. Spatial and temporal soliton pulse generation by transmission of chaotic signals using fiber optic link. J. Optics Res., 16: 121-123.

Amiri, I.S., A. Afroozeh and M. Bahadoran, 2011. Simulation and analysis of multisoliton generation using a PANDA ring resonator system. Chinese Phys. Lett., Vol. 28.

Amiri, I.S., S.E. Alavi, M. Bahadoran, A. Afroozeh and H. Ahmad, 2015. Nanometer bandwidth soliton generation and experimental transmission within nonlinear fiber optics using an add-drop filter system. J. Comput. Theor. Nanosci., 12: 221-225.

Bahadoran, M., A. Afroozeh, J. Ali and P.P. Yupapin, 2012. Slow light generation using microring resonators for optical buffer application. Opt. Eng., 51: 044601-1-044601-8.

Gomez, N., L.F. Mingo, J. Bobadilla, F. Serradilla and J.A.C. Manzano, 2010. Particle swarm optimization models applied to neural networks using the R language. WSEAS. Trans. Syst., 9: 192-202.

Hu, X., Y. Shi and R. Eberhart, 2004. Recent advances in particle swarm. Proceedings of the 2004 Congress on Evolutionary Computation (CEC'04) Vol. 1, June 19-23, 2004, IEEE, Portland, Oregon, ISBN:0-7803-8515-2, pp: 90-97.

Jalil, M.A., A. Abdolkarim, T. Saktioto, C.T. Ong and P.P. Yupapin, 2012. Generation of THz frequency using PANDA ring resonator for THz imaging. Int. J. Nanomed., 7: 773-779.

Kennedy, J., 1997. The particle swarm: Social adaptation of knowledge. Proceedings of the 1997 IEEE International Conference on Evolutionary Computation, April 13-16, 1997, IEEE, Indianapolis, Indiana, USA., ISBN:0-7803-3949-5, pp: 303-308.

Meila, M., 2006. The uniqueness of a good optimum for k-means. Proceedings of the 23rd International Conference on Machine Learning, June 25-29, 2006, ACM, Pittsburgh, Pennsylvania, ISBN:1-59593-383-2, pp: 625-632.

Merwe, V.D.D.W. and A.P. Engelbrecht, 2003. Data clustering using particle swarm optimization. Proceedings of the 2003 Congress on Evolutionary Computation (CEC'03) Vol. 1, December 8-12, 2003, IEEE, Canberra, Australia, ISBN:0-7803-7804-0, pp: 215-220.

Omran, M., A. Salman and A.P. Engelbrecht, 2002. Image classification using particle swarm optimization. Proc. Asia Pacific Conf. Simulated Evol. Learn., 1: 18-22.

Premalatha, K. and A.M. Natarajan, 2010. Hybrid PSO and GA models for document clustering. Intl. J. Adv. Soft Comput. Appl., 2: 302-320.

Robinson, J. and Y. Rahmat-Samii, 2004. Particle swarm optimization in electromagnetics. IEEE. Trans. Antennas Propag., 52: 397-407.

Sadeghierad, M., A. Darabi, H. Lesani and H. Monsef, 2010. Optimal design of the generator of microturbine using genetic algorithm and PSO. Int. J. Elect. Power Energy Syst., 32: 804-808.

Shakerian, R., S.H. Kamali, M. Hedayati and M. Alipour, 2011. Comparative study of ant colony optimization and particle swarm optimization for grid scheduling. J. Math. Comput. Sci., 2: 469-474.

Shi, Y. and R. Eberhart, 1998. A modified particle swarm optimizer. Proceedings of the 1998 IEEE International Conference on Evolutionary Computation and the IEEE World Congress on Computational Intelligence, May 4-9, 1998, IEEE, Anchorage, Alaska, USA., ISBN:0-7803-4869-9, pp: 69-73.

Tsai, C.Y. and I.W. Kao, 2011. Particle swarm optimization with selective particle regeneration for data clustering. Expert Syst. Applic., 38: 6565-6576.

Urade, H.S. and R. Patel, 2012. Performance evaluation of dynamic particle swarm optimization. Intl. J. Comput. Sci. Netw., 1: 1-14.