International Journal of Innovative
Computing, Information and Control
Volume **6**, Number **7**, July **2010**

ICIC International ©2010 ISSN 1349-4198
pp. 3177–3192

# A NEW HYBRID ALGORITHM BASED ON PSO, SA, AND K-MEANS FOR CLUSTER ANALYSIS

Bahman Bahmani Firouzi[1], Mokhtar Sha Sadeghi[2] and Taher Niknam[2]

[1]Islamic Azad University, Marvdasht Branch
Marvdasht, Iran
bahman_bah@yahoo.com

[2]Electronic and Electrical Engineering Department
Shiraz University of Technology
Modares Blvd., Shiraz, Iran
{ niknam; shasadeghi }@sutech.ac.ir

Received January 2009; revised June 2009

Abstract. *It is well known that k-means algorithm is one of the most widely used clustering techniques. However, solutions of k-means algorithm depend on the initialization of cluster centers and the final solution converges to local minima. This paper presents a hybrid evolutionary algorithm to overcome k-means algorithm shortcomings. The proposed algorithm is elaborately developed by combining Particle Swarm Optimization (PSO), Simulated Annealing (SA), and k-means algorithms called PSO-SA-K that helps find better cluster partition. The efficiency of the proposed clustering technique is evaluated by several benchmark data sets. The simulation results show that the proposed algorithm outperforms the previous approaches such as PSO, SA, combination of PSO and SA (PSO-SA), Ant Colony Optimization (ACO), combination of k-means and PSO (K-PSO), combination of Nelder Mead simplex search and PSO (NM-PSO), K-NM-PSO, Genetic Algorithm (GA), Tabu Search (TS), Honey Bee Mating Optimization (HBMO), and k-means for partitional clustering problem.*

**Keywords:** Data clustering, Hybrid evolutionary optimization algorithm, K-means algorithm, Simulated annealing (SA), Particle swarm optimization (PSO)

1. **Introduction.** In data analysis and pattern classification, categorizing a set of data pointing to some non-overlapping clusters is an important topic. It has many applications such as codebook design, data mining, image segmentation, data compression, etc. Many efficient clustering algorithms have been developed for data sets of different distributions in the past several decades [1-3].

Clustering algorithms are divided into hierarchical and partitional categories. In hierarchical clustering, the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions take place which may run from a single cluster containing all objects to n clusters each of them containing a single object. Hierarchical clustering [4] is subdivided into agglomerative methods which proceed by series of fusions of n objects into groups, and divisive methods which separate n objects successively into finer groupings. Partitional clustering [5], on the other hand, attempts to directly decompose the data set into a set of disjoint clusters. The criterion function that the clustering algorithm tries to minimize may emphasize the local structure of the data, as by assigning clusters to peaks in the probability density function, or the global structure. Typically the global criteria involve minimizing some measures of dissimilarities in the samples within each cluster, while maximizing the dissimilarity of different clusters [5-7]. K-means algorithm is the