Evaluation of the Potential of Swarm Clustering Techniques in Hyperspectral Data

Farhad Samadzadegan, Ahmad Abootalebi, Amin Alizadeh

Department of geomatics engineering, Faculty of engineering, University of Tehran
North Amirabad, Tehran, Iran
(samadz, abootalebi, a.alizadeh) @ ut.ac.ir

Abstract. Hyperspectral imaging has become a fast growing technology in remote sensing due to recent advances of hyperspectral imaging sensors. With high spectral resolution of current hyperspectral imaging systems, many subtle objects and materials can now be discovered and extracted. A fundamental step in the transformation of the hyperspectral data into objects is a segmentation of hyperspectral data through a clustering process. Nevertheless, due to high dimensionality of these data, the performance of most of traditional clustering methods such as k-means, significantly decrease and in most of cases these techniques trap in local optimum solution. To overcome the problems of traditional clustering method, a method based on Particle Swarm Optimization (PSO) in clustering of hyperspectral imagery is presented in this paper. PSO algorithm is a population based search algorithm on the simulation of the social behavior of birds whit-in a flock. This paper evaluates the potential of two swarm clustering techniques in hyperspectral data. First is based on PSO and second is a hybrid algorithm based on integration of PSO and k-means (PSOKM). The PSOKM algorithm not only helps the KM clustering escape from local optima but also overcomes the shortcoming of the slow convergence speed of the PSO algorithm.

Keywords: Clustering, hyperspectral data, K-means, Particle Swarm Optimization

1 Introduction

From a methodological viewpoint, a classification process consists of associating a pattern (sample) to a class label randomly chosen from a predefined set of class labels. In the literature, two main approaches to the classification problem have been proposed: 1) the supervised approach and 2) the unsupervised approach. Supervised techniques require the availability of a training set for learning the classifier. Unsupervised methods, known also as clustering methods, perform classification just by exploiting information conveyed by the data, without requiring any training sample set. The supervised methods offer higher classification accuracy compared to the unsupervised ones, but in some applications, it is necessary to resort to unsupervised techniques because training information is not available. In this paper, we focus the attention on hyperspectral image clustering. Compared with conventional multispectral data, hyperspectral data are characterized by a higher spectral resolution, thus giving the opportunity to push further the information extraction capability. However, hyperspectral data involves a greater quantity of data to memorize and to process [1]. K-means is one of the most popular clustering algorithms for handling massive datasets. This algorithm is efficient at clustering large data sets because its computational complexity only grows linearly with the number of data points [2]. Unfortunately, k-means algorithm may converge to solutions that are not optimal. This paper presents a Particle Swarm Optimization (PSO) clustering algorithm for overcoming the existing problems of traditional k-means.

2 Basic concepts in Data Clustering

Historically, the notion of finding useful patterns in data has been given a variety of names including data clustering, data mining, knowledge discovery, pattern recognition, information extraction, etc [2]. Data clustering is an analytic process designed to explore data by discovering of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. Data clustering is a difficult problem in remote sensing data as the clusters in data may have different shapes and sizes. In the background of clustering techniques, the following terms are used in this paper [4]; A pattern (or feature vector), z, is a single object or data point used by the clustering algorithm; A feature (or attribute) is an individual component of a pattern; A cluster is a set of similar patterns, and patterns from different clusters are not similar; A distance measure is a metric used to evaluate the similarity of patterns. The clustering problem can be formally defined as follows[4] : Given a data set $Z=\{z_1,z_2, \ldots ,z_p, \ldots ,z_{Np}\}$ where z_p is a pattern in the N_d-dimensional feature space, and N_p is the number of patterns in Z, then the clustering of Z is the partitioning of Z into K clusters $\{C_1,C_2, \ldots ,C_K\}$ satisfying the following conditions:

- Each pattern should be assigned to a cluster, i.e. $\cup_{j=1}^{k} C_j = Z$
- Each cluster has at least one pattern assigned to it, i.e. $C_k \neq 0,\ k = 1, \ldots, K$
- Each pattern is assigned to one and only one cluster $C_k \cap C_j = 0, \text{where } k \neq j$

Clustering is the process of identifying natural groupings within multidimensional data based on feature space, similarity measures are fundamental components in most clustering algorithms [4]. The most popular way to evaluate a similarity measure is the use of distance measures. The most widely used distance measure is the Euclidean distance, defined as:

$$d(z_i,z_j) = \sqrt{\sum_{k=1}^{N_d}(z_{i,k} - z_{j,k})^2} = \|z_{i-}z_j\| \qquad \textbf{(1)}$$

Generally, clustering algorithms can be categorized into partitioning methods, hierarchical methods, density-based methods, grid-based methods, and model-based methods. An excellent survey of clustering techniques can be found in [6]. Partitional clustering algorithms divide the data set into a specified number of clusters and then evaluate them by some criteria. These algorithms try to minimize certain criteria (e.g. a square error function) and can therefore be treated as optimization problems [9].
The most widely used partitional algorithm in clustering techniques is the iterative k-means approach [6]. The objective optimizing J is the k-means algorithm:

$$J_{K-\text{means}} = \sum_{j=1}^{K} \sum_{\forall z_p \in C_k} d^2(z_p,m_k) \qquad (2)$$

Where m_k is the centroid of the k-th cluster. The membership and weight functions u for k-means are defined as:

$$u(m_k|z_p) = \begin{cases} 1 & \text{if } d^2(z_p,m_k) = \arg\min_k\{d^2(z_p,m_k)\} \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

Consequently, the k-means method minimizes the intra-cluster distance. The k-means algorithm starts with k centroids (initial values are randomly selected or derived from a priori information). Then, each pattern z_p in the data set is assigned to the closest cluster (i.e. closest centroid). Finally, the centroids are recalculated according to the associated patterns. This procedure is repeated until convergence is achieved [11].

3 Particle Swarm Optimization

PSO is a population based stochastic optimization technique inspired by the social behavior of bird flock (and fish school, etc.), as developed by Kennedy and Eberhart [5]. As a relatively new evolutionary paradigm, it has grown in the past decade and many studies related to PSO have been published [3]. In PSO, each particle is an individual, and the swarm is composed of particles. The problem solution space is formulated as a search space. Each position in the search space is a correlated solution of the problem. Particles cooperate to find the best position (best solution) in the search space (solution space). Each particle moves according to its velocity which is computed as:

$$v_i(t+1)=v_i(t)+c_1r_1(pbest_i(t) - x_i(t))+c_2r_2(gbest(t) - x_i(t)) \qquad (4)$$

$$x_i(t+1)=x_i(t)+v_i(t+1) \qquad (5)$$

In equation (4) and (5), $x_i(t)$ is the position of particle i at time t, $v_i(t)$ is the velocity of particle i at time t, $pbest_i(t)$ is the best position found by particle i itself so far, gbest(t) is the best position found by the whole swarm so far, ω is an inertia weight scaling the previous time step velocity, c_1 and c_2 are two acceleration coefficients that scale the influence of the best personal position of the particle ($pbest_i(t)$) and the best global position (gbest(t)), r_1 and r_2 are random variables between 0 and 1 [13].

4 Clustering of Hyperspectral Data Using Particle Swarm Optimization Algorithm

Recently a family of nature inspired algorithms, known as Swarm Intelligence (SI), has attracted several researchers from the field of pattern recognition and clustering. Clustering techniques based on the SI tools have reportedly outperformed many classical methods of partitioning a complex real world dataset. Swarm Intelligence is a relatively new interdisciplinary field of research, which has gained huge popularity in these days. Algorithms belonging to the domain, draw inspiration from the collective intelligence emerging from the behavior of a group of social insects (like bees, termites and wasps). When acting as a community, these insects even with very limited individual capability can jointly (cooperatively) perform many complex tasks necessary for their survival. Problems like finding and storing foods, selecting and picking up materials for future usage require a detailed planning, and are solved by insect colonies without any kind of supervisor or controller. Particle Swarm Optimization (PSO) [5] is another very popular SI algorithm for global optimization over continuous search spaces. Since its advent in 1995, PSO has attracted the attention of several researchers all over the world resulting into a huge number of variants of the basic algorithm as well as many parameter automation strategies [1]. For clustering data there are two categories, PSO and PSOKM that they are considered in the following sections.

4.1 PSO Clustering

Research efforts have made it possible to view data clustering as an optimization problem. This view offers us a chance to apply PSO algorithm for evolving a set of candidate cluster centroids and thus determining a near optimal partitioning of the dataset at hand. An important advantage of the PSO is its ability to cope with local optima by maintaining, recombining and comparing several candidate solutions simultaneously. In contrast, local search heuristics, such as the simulated annealing algorithm [12], only refine a single candidate solution and are notoriously weak in coping with local optima. Deterministic local search, which is used in algorithms like the K-means, always converges to the nearest local optimum from the starting position of the search [1].
In the context of clustering, a single particle represents the N_c cluster centroid vectors. That is, each particle x_i is constructed as follows:

$$x_i = (m_{i1}, \dots, m_{ij}, \dots, m_{in_c}) \qquad (6)$$

Where m_{ij} refers to the j-th cluster centroid vector of the i-th particle in a cluster. Therefore, a swarm represents a number of candidate clusters for the current data vectors.
The population-based search of the PSO algorithm reduces the effect of initial conditions, as opposed to the K-means algorithm; the search starts from multiple positions in parallel.

$$F = \frac{d_{max} + j_e}{d_{min}} \qquad (7)$$

Where d_{max}, d_{min}, j_e and z_{max} like below are defined: $j_e = \frac{\sum_{j=1}^{N_c} d(z_p, m_{ij})/|c_j|}{N_c}$; $d_{max}(z_i, x_i) = \max\left\{\frac{d(z_p, m_{ij})}{|c_{ij}|}\right\}$ and $d_{min}(x_i) = \min\{d(m_{ij}, m_{ik})\}$.

Using the standard gbest PSO, data vectors can be clustered as follows:

1. Initialize each particle to contain N_c , randomly selected cluster centroids. For example Iris data set has four dimension and three clusters.
2. For t = 1 to t_{max}

(a) For each particle i
(b) For each data vector z_p
i) Calculate the Euclidean distance $d(z_p,m_{ij})$ to all cluster centroids
ii) Assign z_p to cluster c_{ij} such that $d(z_p, m_{ij}) = \min \forall c = 1, \ldots, N_c$
iii) Calculate the fitness using equation (7)
(c) Update the global best and local best positions
(d) Update the cluster centroids using equations (4) and (5)

Where t_{max} is the maximum number of iterations [8].

4.2 Hybrid PSOKM Algorithm for Clustering

The PSOKM is an optimization algorithm combining the PSO with the K-means, in order to solve the problem of data clustering. Similar to the GA, the PSO algorithm is a global algorithm, which has a strong ability to find global optimistic result; this PSO algorithm, however, has a disadvantage that the search around global optimum is very slow. The K-means algorithm, on the contrary, has a strong ability to find local optimistic result for clustering problem, but its ability to find the global optimistic result is weak. By combining the PSO with the K-means, a new algorithm referred to as PSOKM hybrid algorithm is formulated. After the encoding of the string of the particles, the execution of PSOKM is as follow:

- **Step 1)** Initialize position vector of particle and associated velocity of all particles in the population randomly.
- **Step 2)** Evaluate the fitness function for each particle. We use metric function which proposes by (7)
- **Step 3)** Compare every particle's fitness value with previous particle's best solution (pbest). If current solution is better than a previous value (pbest), then update pbest with current solution.
- **Step 4)** Compare fitness evaluation with the population's overall previous best. If current value is better than the gbest (the global version of the best value), then reset gbest to the current particle's value and position.
- **Step 5)** Use the one step of K-means algorithm to replace the result of the gbest. The cluster centres encoded in the gbest are replaced by the mean points of the respective clusters (8):

$$z_j = \frac{1}{N_j} \sum_{x_i \in C_j} x_i \qquad (8)$$

  where N_j is the number of points belonging to cluster C_j. The effect of the K-means algorithm is to direct the best solution towards the area of the training data. The drawback of the hybridization is that the running time considerably grows as the number of K-means step increases. For better convergence and lower computing time purpose, the Step 5 work in the initial five iterations (or less) is enough.
- **Step 6)** Change velocities and positions with Eq. (4) and Eq. (5).
- **Step 7)** Repeat from Step2 to Step6 until the predefined number of iterations is completed [14].

5 Experimental Investigations

5.1 Dataset

In order to evaluate the PSO method, a sub-image of AVIRIS data [16] with the size of 145×145 pixels was used. In addition, [17] was used for implementing and evaluating the algorithm. This image was taken over the northwest Indiana's Indian Pine test site in June 1992 and has sixteen classes. The data has 220 spectral bands, about 10 nm apart between 0.4 to 2.45 µm with a spatial resolution of 20 m. The twenty water absorption bands (numbered 104-108, 150-163, and 220) were removed from the original image. In addition, fifteen noisy bands 1-3, 103, 109-112, 148-149, 164-165, and 217-219 as observed from visual inspection , were also removed, resulting in a total of 185 bands. The number of labelled samples per class is given in Table 1. The ground truth map is shown in Figure 4. Since some classes are too small to retain enough disjoint samples for training and testing, seven classes were neglected, leaving nine classes for the experiments which among classes in this paper have used four classes with name: Corn, Grass/pasture, Hay-windrowed and soybean-clean[7].

Table 1. Cluster Size of AVIRIS Data.

Class name	Numbers of sample	Class name	Numbers of sample
Corn-notill	1434	Soybeans-notill	968
Corn-min	834	Soybeans-min	2468
Grass/Pasture	497	Soybeans-clean	614
Grass/trees	747	Woods	1294
Hay-windrowed	489		

Fig 1. a) Ground truth of the area with 16 classes. b) Color composite of the image subset

The k-means and particle swarm optimization were developed based on the parameters listed in Table 2. Note that parameters of PSOKM are similar to PSO.

Table 2. Parameters used in the clustering of hyperspectral datasets

Algorithm	Parameters	Value
k-means	Maximum number of iterations	100
Particle swarm optimization algorithm	Number of particles	10
	Inertia weight	0.4
	Acceleration constant 1	1
	Acceleration constant 2	1
	Maximum number of iteration	100

5.2 Accuracy Assessment

In this paper, confusion matrix used to evaluate the true labels and the labels returned by the clustering algorithms as the quality assessment measure [15]. From the confusion matrix we calculate the Kappa Coefficient. The kappa coefficient uses all of the information in the confusion matrix in order for the chance allocation of labels to be taken into consideration. The kappa coefficient is defined in (9) also for individual classes, the Khat index is calculated using the formula in (10) that average of numbers of any class is average accuracy (AA).

$$\hat{k} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} \times x_{+i})} \qquad (9)$$

$$k_i = \frac{N x_{ii} - x_{i+} \times x_{+i}}{N \times x_{i+} - x_{i+} \times x_{+i}} \qquad (10)$$

In equation (9), $\hat{k}$ is the kappa coefficient and in equation (10) k_i is the Khat index for individual classes, r is the number of columns (and rows) in a confusion matrix, x_{ii} is entry (i, i) of the confusion

matrix, x_{i+} and x_{+i} are the marginal totals of row i and column j, respectively, and N is the total number of observations. Table 3. shows the matrix and kappa coefficient and overall accuracy of k-means, PSO and PSOKM in AVIRIS dataset.
By comparing the counts in each class, a striking difference to the particle swarm optimization result is clearly observed. For the two classes of major interest in this study, the corn-min class and soybean-clean class, the differences between potential of k-means and PSO based methods are quite significant (Figure 2).

Table 3. Confusion matrix and kappa coefficient of k-means and particle swarm optimization algorithms in the northwest Indiana's Indian.

		Reference Data				
		Soybeans-clean	Grass/Pasture	Corn-min	Hay-windrowed	Total
k-means	Soybeans-clean	290	0	300	24	614
	Grass/Pasture	18	209	10	260	497
	Corn-min	526	15	268	25	834
	Hay-windrowed	39	0	0	450	489
	Total	873	224	578	759	2434
	Kappa coefficient = 0.3304					
	Overall accuracy=0.5					
PSO		Reference Data				
		Soybeans-clean	Grass/Pasture	Corn-min	Hay-windrowed	Total
	Soybeans-clean	313	0	199	102	614
	Grass/Pasture	1	328	7	161	497
	Corn-min	257	0	550	27	834
	Hay-windrowed	0	0	0	489	489
	Total	339	328	988	779	2434
	Kappa coefficient = 0.5828					
	Overall accuracy=0.6902					
PSOKM		Reference Data				
		Soybeans-clean	Grass/Pasture	Corn-min	Hay-windrowed	Total
	Soybeans-clean	313	0	288	13	614
	Grass/Pasture	9	328	6	154	497
	Corn-min	247	0	580	7	834
	Hay-windrowed	0	0	0	489	489
	Total	679	328	764	663	2434
	Kappa coefficient = 0.5957					
	Overall accuracy=0.7025					

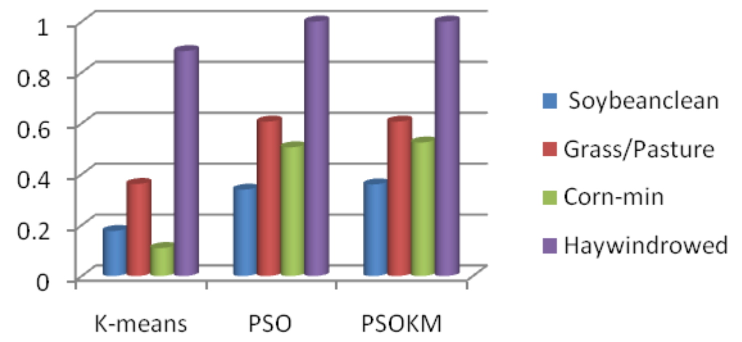

Fig 2. Kappa coefficient for any cluster in different methods

6 Conclusion

This paper evaluated the potential of two PSO based clustering methods for extracting four classes with names: Soybeans, Hay-windrowed, Grass/pasture and Corn-min from AVIRIS imagery. First method employs the particle swarm optimization algorithm to search for the set of cluster centers that minimizes a given clustering metric. Although this method shows a significant improvement in finding the optimal cluster centers, it has a slow rate of convergency to achieve the global optima. The second method, PSOKM algorithm, not only improved the convergence speed of PSO but also helps K-means to escape from local optima.

However, the main point that should be considered in further investigation is determination of the optimum value of PSO's parameters. These parameters have direct effect on the potential of PSO based clustering methods.

Reference

[1] Abraham A, Das S, Roy S. Swarm Intelligence Algorithms for Data Clustering.

[2] Ajith A, Crina G, Vitorino R. Swarm Intelligence in Data Mining, Studies in Computational Intelligence (series), Vol. 34, Springer-Verlag, ISBN: 3-540-34955-3, 267 p., Hardcover. (Eds.) (2006).

[3] Ali M.M, Kaelo P. Improved Particle Swarm Algorithms for Global Optimization. Applied Mathematics and Computation 196, 578–593 (2008).

[4] Jain A.K, Murty M.N, Flynn P.J, (1999). Data Clustering: A Review, ACM Computing Surveys, Vol. 31, No. 3, Sep. 1999.

[5] Kennedy J, Eberhart R.C. Particle Swarm Optimization. In: Proceedings of the IEEE intenational conference on neural networks, pp. 1942–1948. IEEE Press, New Jersey (1995).

[6] Kotsiantis S, Pintelas P. Recent advances in clustering: a brief survey, WSEAS Transactions on Information Science and Applications 1:73-81. (2004).

[7] Mojaradia B, Emami H, Varshosaz M, Jamali S. A NOVEL BAND SELECTION METHOD FOR HYPERSPECTRAL DATA ANALYSIS , The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B7. Beijing 2008.

[8] Murthy J, Reddy P. A Comparative Analysis of Unsupervised K-means, PSO and Self-Organizing PSO for Image Clustering, IEEE, 2007.

[9] Omran M, Salman A. Engelbrecht. AP,Image Classification using Particle Swarm Optimization. In Conf. on Simulated Evolution and Learning, vol. 1, pp 370–374, 2002.

[10] Paoli A, Melgani F, Passoli E. clustering of hyperspectral images based on multi objective particle swarm optimization, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING 1,2007.

[11] Saeedi S.F, Samadzadegan N. El-sheimy,object extraction from lidar data using an artificial swarm bee colony clustering algorithm, ISPRS, 2009.

[12] Selim S.Z, Alsultan. K. (1991), A simulated annealing algorithm for the clustering problem. Pattern recognition, 24(7), pp. 1003-1008.

[13] Yang F, Zhang C, Sun T. Particle Swarm Optimization and Differential Evolution in Fuzzy Clustering, Springer-Verlag Berlin Heidelberg 2009.

[14] Ye F, Yi C Chen, Alternative KPSO-Clustering Algorithm, Tamkang Journal of Science and Engineering, Vol. 8, No 2, pp. 165_174 (2005) 165.

[15] Zhong S, Ghosh J. A comparative study of generative models for document clustering; in SIAM Int. Conf. Data Mining Workshop on Clustering High Dimensional Data and Its Applications, San Francisco, CA, May 2003.

[16] Landgrebe, D., ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/ (accessed 10 Sep. 2009).

[17] Varshney P.K, Arora M.K. Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data. *Springer*, Berlin, 2004.