A Hybrid Rough Set – Particle Swarm Algorithm for Image Pixel Classification

Swagatam Das[1], Ajith Abraham[2] and Subir Kumar Sarkar[1]
[1]*Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata 700032, India*
[2]*IITA Professorship Program, School of Computer Science, Yonsei University, Seoul, Korea*
ajith.abraham@ieee.org

Abstract

This article presents a framework to hybridize the rough set theory with a famous swarm intelligence algorithm known as Particle Swarm Optimization (PSO). The hybrid rough-PSO technique has been used for grouping the pixels of an image in its intensity space. Medical and remote sensing satellite images become corrupted with noise very often. Fast and efficient segmentation of such noisy images (which is essential for their further interpretation in many cases) has remained a challenging problem for years. In this work, we treat image segmentation as a clustering problem. Each cluster is modeled with a rough set. PSO is employed to tune the threshold and relative importance of upper and lower approximations of the rough sets. Davies–Bouldin clustering validity index is used as the fitness function, which is minimized while arriving at an optimal partitioning.

1. Introduction

Image segmentation may be defined as the process of dividing an image into disjoint homogeneous regions. These homogeneous regions usually contain similar objects of interest or part of them. The extent of homogeneity of the segmented regions can be measured using some image property (e. g. pixel intensity [1]). On the other hand, clustering can be defined as the optimal partitioning of a given set of n data points into c subgroups, such that data points belonging to the same group are as similar to each other as possible whereas data points from two different groups share the maximum difference.

Image segmentation can be treated as a clustering problem where the features describing each pixel correspond to a pattern, and each image region (i.e.a segment) corresponds to a cluster [1]. Therefore many clustering algorithms have widely been used to solve the segmentation problem (e.g., K-means [2], FCM [3], ISODATA [4] and Snob [5]).

Popular hard clustering approaches do not consider overlapping of classes that occur in many practical image segmentation problems. For example, in remote sensing satellite images, a pixel corresponds to an area of the land space, which may not necessarily belong to a single type of land cover. This in turn indicates that the pixels in a satellite image can be associated with a large amount of imprecision and uncertainty. Therefore, application of the principles of fuzzy set theory has remained a popular choice for the researchers in this domain [6-7].

However, the rough set theory, pioneered by Pawlak in mid 1980's [8], has emerged as a promising mathematical tool for extracting knowledge from datasets which contain imperfection, such as noise, unknown values or errors due to inaccurate measuring equipment. In this work rough sets are used to model the clusters in terms of upper and lower approximations. PSO [9], which gained huge popularity as a naturally inspired optimization tool in recent times, is used to tune the threshold, and relative importance of upper and lower approximation parameters of the sets. The Davies–Bouldin clustering validity index is used as the fitness function of the PSO, that is minimized. We present comparison of our hybrid algorithm with the classical FCM based segmentation [6] and another state-of-the-art image segmentation technique [10] over some well chosen gray scale images. Such comparisons reflect the superiority of the proposed method. The rest of the paper is organized as follows. Section 2 provides a brief outline of the rough set theory. In section 3 the PSO algorithm and its proposed modifications have been discussed. In section 4, we present the hybrid algorithm for image pixel classification. Results are presented and discussed in section 5 before drawing conclusions in section 6.

2. The Rough Sets

Introduced by Pawlak [8] in the 1980's, rough set theory constitutes a sound basis for discovering patterns in hidden data and thus have extensive applications in data mining in distributed systems. It has recently emerged as a major mathematical tool for managing uncertainty that arises from granularity in the domain of discourse—that is, from the indiscernibility between objects in a set. The intention is to approximate a rough (imprecise) concept in the domain of discourse by a pair of exact concepts, called the lower and upper approximations. These exact concepts are determined by an indiscernibility relation on the domain, which, in turn, may be induced by a given set of attributes ascribed to the objects of the domain. The lower approximation is the set of objects definitely belonging to the vague concept, whereas the upper approximation is the set of objects possibly belonging to the same. Fig. 1 provides a schematic diagram of a rough set.

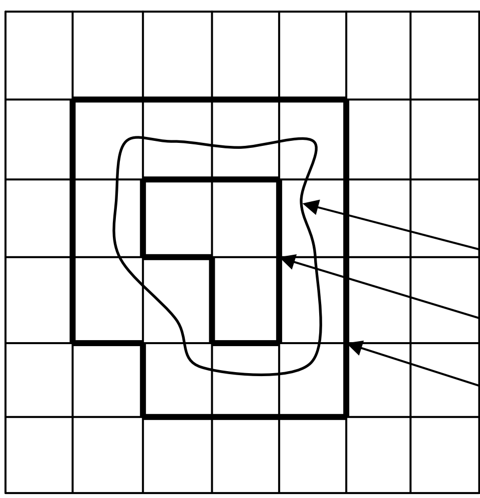

Figure1: The rough boundaries $R_L(A)$-the lower approximation and $R_U(A)$-the upper approximation of a given point set $A \subseteq X$-the universe of discourse.

3. The Particle Swarm Optimization (PSO)

The PSO algorithm, as first described by Eberhart and Kennedy is reminiscent of the behavior of flock of birds or the sociological behavior of a group of people. In PSO [9, 11], a population of particles is initialized with random positions:

$$\vec{Z}_i(t) = [Z_{i,1}(t), Z_{i,2}(t), \ldots Z_{i,d}(t)]$$

and velocities:

$$\vec{v}_i(t) = [v_{i,1}(t), v_{i,2}(t), \ldots v_{i,d}(t)]$$

in d-dimensional space. A fitness function, f is evaluated, using the particle's positional coordinates as input values. Positions and velocities are adjusted, and the function is evaluated with the new coordinates at each time-step. The velocity and position update equations for the p-th dimension of the i-th particle in the swarm may be given as follows:

$$v_{ip}(t+1) = \omega . v_{ip}(t) + C_1 . \varphi_1 . (P_{lip} - Z_{ip}(t)) + C_2 . \varphi_2 . (P_{gp} - Z_{ip}(t))$$
$$Z_{ip}(t+1) = Z_{ip}(t) + v_{ip}(t+1) \quad (1)$$

The variables φ_1 and φ_2 are random positive numbers, drawn from a uniform distribution, and with an upper limit $\varphi_{max,}$ which is a parameter of the system. C_1 and C_2 are called acceleration constants, and ω is the inertia weight. P_{li} is the best solution found so far by an individual particle, while P_g represents the fittest particle found so far in the entire community.

4. The Proposed Algorithm

A pattern is a physical or abstract structure of objects. It is distinguished from others by a collective set of attributes called *features*, which together represent a pattern. Let P ={P_1, P_2, ..., P_n} be a set of n patterns or data points, each having d features. These patterns can also be represented by a profile data matrix $X_{n\times d}$ having n d-dimensional row vectors. The i-th row vector $\vec{X}_i$ characterises the i-th object from the set P and each element $X_{i,j}$ in $\vec{X}_i$ corresponds to the j-th real value feature (j = 1, 2, ...,d) of the i-th pattern (i =1,2,...., n). Given such an $X_{n\times d}$, a partitional clustering aalgorithm tries to find a partition C = {C_1, C_2,......, C_c} such that the similarity of the patterns in the same cluster C_i is maximum and patterns from different clusters differ as far as possible. The partitions should maintain the following properties:

1) $C_i \neq \emptyset \quad \forall i \in \{1,2,...,c\}$
2) $C_i \cap C_j = \emptyset, \quad \forall_i \neq j$ and $i, j \in \{1,2,...,c\}$
3) $\bigcup_{i=1}^{c} C_i = P$

4.1 The Rough c-means Algorithm

In rough c-means (RCM) algorithm, the concept of c-means clustering [12] is extended by viewing each cluster as an interval or rough set [13]. A rough set Y is characterized by its lower and upper approximations R_L (Y) and R_U (Y) respectively. This permits overlaps between clusters. Here an object $\vec{X}_i$ can be part of at most one lower approximation. If $\vec{X}_k \in R_1(Y)$ of cluster Y , then simultaneously $\vec{X}_k \in R_2(Y)$. If $\vec{X}_i$ is not a part of any lower approximation, then it belongs to two or more upper approximations. Here the cluster center $\vec{Z}_i$ of cluster C_i is computed as:

$$\vec{Z}_i = w_{low} \frac{\sum_{\vec{X}_k \in R_1(Y)} \vec{X}_k}{|R_1(Y)|} + w_{up} \frac{\sum_{\vec{X}_k \in [R_2(Y)-R_i(Y)]} \vec{X}_k}{|R_2(Y)-R_1(Y)|}$$

if $R_2(A) - R_1(A) \neq \phi$

$$= w_{low} \frac{\sum_{\vec{X}_k \in R_1(Y)} \vec{X}_k}{|R_1(Y)|} \quad \text{otherwise.} \qquad (2)$$

where the parameters w_{low} and w_{up} correspond to the relative importance of the lower and upper approximations respectively. Here $|R_1(Y)|$ indicates the number of pattern points in the lower approximation of cluster Y, while $|R_2(Y)-R_1(Y)|$is the number of elements in the rough boundary lying between the two approximations. In RCM (Rough c-means), a threshhold parameter needs special mention. If the difference of distances (Euclidean usually) of an object $\vec{X}_k$ from two cluster centers $\vec{Z}_i$ and $\vec{Z}_j$ of clusters C_i and C_j respectively, is lesser than some threshold δ, then $\vec{X}_k \in R_2(C_j)$ and $\vec{X}_k \in R_2(C_i)$ and $\vec{X}_k$ cannot be a member of any lower approximation. Else, $\vec{X}_k \in R_1(C_j)$ such that distance $d(\vec{X}_k, \vec{Z}_i)$ is minimum over the c clusters. It is to be noted that a major disadvantage of the rough c-means algorithm is the involvement of too many user-defined parameters.

4. 2 Effects of Parameters on RCM

It is observed that the performance of the algorithm is dependent on the choice of w_{low}, w_{up} and threshold δ. We allowed $w_{up} = 1 - w_{low}$, $0.5 < w_{low} < 1$ and $0 < \delta < 0.5$. It is to be noted that the parameter threshold measures the relative distance of an object $\vec{X}_k$ from a pair of clusters having centroids $\vec{Z}_i$ and $\vec{Z}_j$. The smaller the value of threshold, the more likely is Xk to lie within the rough boundary (between upper and lower approximations) of a cluster. This implies that only those points which definitely belong to a cluster (lie close to the centroid) occur within the lower approximation. A large value of threshold implies a relaxation of this criterion, such that more patterns are allowed to belong to any of the lower approximations. The parameter wlow controls the importance of the objects lying within the lower approximation of a cluster in determining its centroid. A lower w_{low} implies a higher w_{up}, and hence an increased importance of patterns located in the rough boundary of a cluster towards the positioning of its centroid.

4. 3 Tuning the Parameters with PSO

In this work we employed a PSO algorithm to determine the optimal values of the parameters w_{low} and δ for each c (number of clusters). As the fitness function of the PSO, we have chosen a statistical-mathematical function, also called a cluster validity index, well known as Davies-Bouldin (DB) index [14]. This measure is a function of the ratio of the sum of within-cluster scatter to between-cluster separation, and it uses both the clusters and their sample means. First, we define the *within i-th cluster scatter* and the *between i-th and j-th cluster distance* respectively as,

$$S_{i,q} = \left[\frac{1}{N_i} \sum_{\vec{X} \in C_i} \left\| \vec{X} - \vec{Z}_i \right\|_2^q \right]^{1/q} \qquad (3)$$

$$d_{ij,t} = \left\{ \sum_{p=1}^{d} \left| Z_{i,p} - Z_{j,p} \right|^t \right\}^{1/t} = \left\| \vec{Z}_i - \vec{Z}_j \right\|_t \qquad (4)$$

where q, t ≥ 1, *q* is an integer and *q, t* can be selected independently. N_i is the number of elements in the i-th cluster C_i. Next $R_{i,qt}$ is defined as,

$$R_{i,qt} = \max_{j \in K, j \neq i} \left\{ \frac{S_{i,q} + S_{j,q}}{d_{ij,t}} \right\} \qquad (5)$$

Finally, we define the DB measure as,

$$DB(c) = \frac{1}{c} \sum_{i=1}^{c} R_{i,qt} \qquad (6)$$

The smallest DB(c) indicates a valid optimal partition.

4. 4 Putting it Altogether

We treat all the pixels of an input image as datapoints. The gray-scale intensity of each pixel serve as a single feature. . Hence, although the data points are single dimensional, the number of data-items is as high as 65,536 for a 256×256 gray image. Then we run a rough c-means algorithm on the image pixel data. Parameters of the RCM are evolved on the run by employing a PSO algorithm. We find that this results into an excellent image segmentation algorithm, which has two special advantages: 1) it removes noisy spots, and it is less sensitive to noise than other techniques. 2)

It yields regions, more homogeneous than the existing methods even in presence of noise.

The major steps of the algorithm can be described as follows:

step 1: Choose the initial means $\vec{Z}_i$ for the c clusters.

step 2: Initialize the population of particles encoding parameters threshold and w_{low}.

step3: Tune the parameters by minimizing the DB index [expression (6)] as the fitness function for the PSO, considering objects lying within the lower approximation of each cluster.

step 4: Assign each data object $\vec{X}_k$ to the lower approximation $R_1(Y)$ or upper approximation $R_2(Y)$ of cluster C_i, by computing the difference in its distance $d(\vec{X}_k, \vec{Z}_i) - d(\vec{X}_k, \vec{Z}_j)$ from cluster centroid pairs $\vec{Z}_i$ and $\vec{Z}_j$.

step 5: If the difference of distances of an object $\vec{X}_k$ from two cluster centers $\vec{Z}_i$ and $\vec{Z}_j$ is lesser than some threshold δ, then $\vec{X}_k \in R_2(C_j)$ and $\vec{X}_k \in R_2(C_i)$ and $\vec{X}_k$ cannot be a member of any lower approximation. Else, $\vec{X}_k \in R_1(C_j)$ such that distance $d(\vec{X}_k, \vec{Z}_j)$ is minimum over the c clusters.

step6 : Compute new mean for each cluster C_i using expression (2).

step 7: Repeat Steps (iii)–(vi) until convergence.

5. Results

We report here the experiments conducted on a test suite of two grayscale images using the rough-PSO hybrid algorithm (for the lack of space we can not report the full set of experiments conducted in this study). In our test-bed, 'robot' comes in 256×256 pixels, while 'the IRS (Indian Remote Sensing Satellite) image of Mumbai is of size 512×512. The IRS image of Mumbai was obtained using the LISS-II sensor. It is available in five bands, viz. blue, green, red and near infra-red. Fig. 5(a) shows the IRS image of a part of Mumbai in the near infrared band. The results obtained using a recent evolutionary fuzzy segmentation algorithm known as FVGA [10] has also been reported for comparison. Table 1 lists the value of DB index (and the corresponding number of clusters) calculated over the final solution in each case. The final result comes as a mean of 25 independent runs of each algorithm. (FVGA and rough-PSO continued up to 50,000 fitness evaluations in each run.

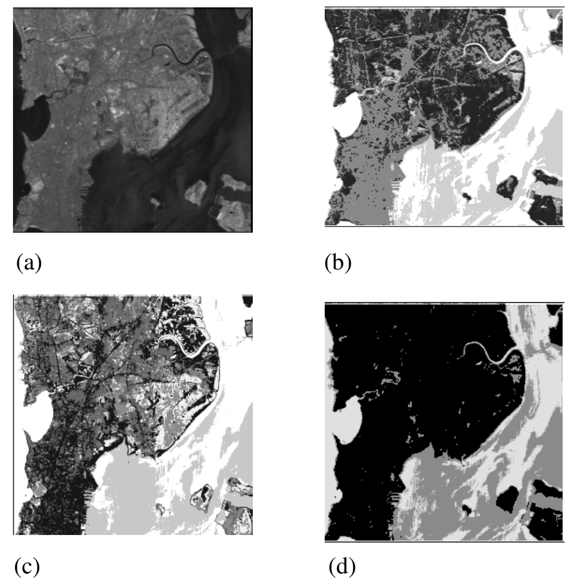

Fig. 2: (a) The original IRS image of Mumbai. (b) Segmentation by FVGA (c) Segmentation by rough-PSO (d) Segmentation with FCM

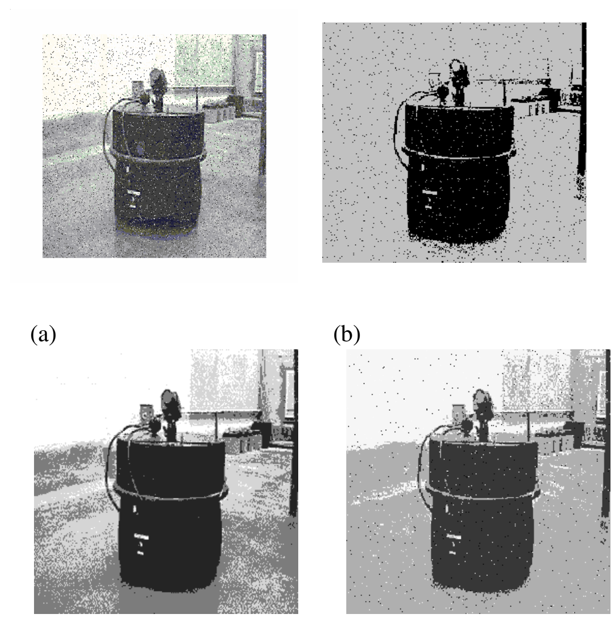

Fig 3: (a) The original Nomadic Super Scout II Robot image (corrupted with salt & pepper noise) (b) Segmentation by FVGA (c) Segmentation by rough-PSO (d) Segmentation with FCM

Table1. Segmentation results for two real life grayscale images (over 25 runs; each run continued up to 50,000 FE)

Image	Number of Classes	Mean and standard deviation of the DB index over the final clustering results of 25 independent runs		
		MEPSO	FVGA	FCM
IRS image of Mumbai	c = 6	**0.7283 (0.0001)**	0.7902 (0.0948)	0.7937 (0.0013)
		2.6631 (0.0018)	2.1193 (0.0826)	2.1085 (0.0043)
The Nomadic Super Scout II Robot	c = 4	**0.2261 (0.0017)**	0.2919 (0.0583)	0.3002 (0.0452)
		0.1837 (0.0062)	0.1922 (0.0096)	0.1939 (0.0921)

6. Conclusion

This paper has presented a new, hybrid algorithm for clustering of images. It has described the formulation of a PSO-based rough c-means clustering algorithm. The relative importance of the upper and lower approximations and the threshold of the rough clusters are optimized using PSO. The DB clustering validity index is chosen as the fitness function being minimized. Results are provided on a remote sensing image and a noisy gray-scale image. As can be perceived from figure 3. (c), the proposed algorithm is especially effective in detecting the correct segments in the image, despite the presence of many noisy spots.

Acknowledgements

This research was supported by the Ministry of Information and Communication, Korea, supervised by the IITA (Institute of Information Technology Assessment) program.

References

[1] Jain, A.K., Murty M.N., and Flynn P.J. (1999): Data Clustering: A Review, ACM Computing Surveys, Vol 31, No. 3, 264-323.

[2] Tou J.T, and Gonzalez R.C. (1974): Pattern Recognition Principles. London, Addison-Wesley.

[3] Trivedi M. M, Bezdek J. C, Low-level segmentation of aerial images with fuzzy clustering , IEEE Trans.on Systems, Man and Cybernetics, Volume 16 , Issue 4 July, 1986.

[4] G. Ball and D. Hall, A Clustering Technique for Summarizing Multivariate Data, Behavioral Science 12, 153-155 (1967).

[5] Wallace, C.S. and Boulton, D.M. (1968), `An Information Measure for Classification', Computer Journal, Vol. 11, No. 2, 1968, pp. 185-194.

[6] Bezdek, J.C. 1981. Pattern recognition with fuzzy objective function algorithms. New York: Plenum.

[7] L.O. Hall, I.B. Özyurt, J.C. Bezdek, Clustering with a genetically optimized approach, IEEE Trans. Evo Comp. 3 (2)(1999) 103–112.

[8] Pawlak, Z., 1991. Rough Sets, Theoretical Aspects of Reasoning about Data. Kluwer Academic, Dordrecht.

[9] Kennedy, J, Eberhart R. Particle swarm optimization, In Proceedings of IEEE International Conference on Neural Networks, (1995) 1942-1948.

[10] Bandyopadhyay, S., and Maulik, U., (2002) Genetic clustering for automatic evolution of clusters and application to image classification, Pattern Recognition, 35, 1197-1208.

[11] Eberhart, R. C., Shi, Y.: Particle swarm optimization: Developments, applications and resources, In Proceedings of IEEE International Conference on Evolutionary Computation, vol. 1 (2001), 81-86.

[12] Lingras, P., West, C., 2002. Interval set clustering of Web users with rough k-means. Technical Report No. 2002-002, Department of Mathematics and Computer Science, St. Mary's University, Halifax, Canada, 2002.

[13] Tou, J.T., Gonzalez, R.C., 1974. Pattern Recognition Principles. Addison-Wesley, London.

[14] Davies, D.L., Bouldin, D.W., A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence 1 (1979) 224–227.