Natural Computing Techniques for Data Clustering and Image Segmentation

J. G. Souza, J. A. F. Costa
Universidade Federal do Rio Grande do Norte – UFRN – Departamento de Engenharia Elétrica
jgomes@dca.ufrn.br; alfredo@dee.ufrn.br

Abstract

This paper presents innovative ways to solve data clustering and image segmentation using Natural computing, a novel approach to solve real life problems inspired in the life. Evolutionary Computing, which is based on the concepts of the evolutionary biology and individual-to-population adaptation, and Swarm Intelligence, which is inspired in the behavior of individuals that, in group, try to achieve better results for a complex optimization problem, are detailed and very experimental results present a comparison between algorithms' implementations.

1. Introduction

Image segmentation techniques are based on Pattern Recognition and Data Mining. The relation between patterns generates classes, or clusters. Considering the image context, the clusters correspond to some semantic meaning in the image, which is, objects.

Among the many methods for data analysis through clustering and unsupervised image segmentation which includes: Nearest Neighbor Clustering, Fuzzy Clustering, Artificial Neural Networks for Clustering [1] current research efforts have been made in the, not even so new, area of evolutionary techniques.

2. Data Clustering and Image Segmentation

To find clusters in a data set is to find relations amongst unlabeled data. The "relation" means that some data are in some way next to another that they can be grouped. It is found in [1] that the components of a clustering task are:

1. *Pattern representation*, which includes feature extraction and feature selection;
2. A *Distance measure* is used to determine pattern proximity;
3. *Clustering* relates to finding the groups (or, labeling the data) and it can be hard or fuzzy.
4. *Data abstraction* extracts a simple and compact representation of a data set (the centroids are chosen).
5. *Assessment of output* is the process of evaluating the clustering result. Cluster validation techniques are approaches used to quantitatively evaluate the clustering results to find the partitioning that best fits the data [2].

3. Natural computing

Natural computing "is the computational version of the process of extracting ideas from nature to develop computational systems, or using natural materials to perform computation" [3] and it can be classified according to the following sections [4]: a) computing inspired by nature, which includes: Artificial Neural Networks, Evolutionary Computing, Swarm Intelligence, and Artificial Immune Systems; b) Simulation of nature; and c) Computing with natural materials.

4. Clustering using Natural Computing

4.1. Genetic K-Means

Genetic Algorithms have been applied to many function optimization problems and are shown to be good in finding optimal and near optimal solutions [4]. The Genetic K-means Algorithm (GKA), developed by [4] aims to solve the partitional clustering algorithm's problem of finding a partition in a given data, with a number of centroids. The objective function, thus, tries to minimize the Total Within Cluster Variation (TWCV), finding the clustering that has centroids attending concepts of *compactness* (patterns from on cluster are similar to each other and different from patterns in other clusters) and *separation* (the clusters' centroids are well-separated, considering a distance measure as the Euclidean Distance). The main aspects of GKA are [4]: a) *Coding*; b) *Initialization;* c) *Selection*; d) *Mutation;* and e) *K-Means Operator (KMO)*. The termination criteria can be, for example, a maximum number of generations, or a user defined threshold for TWCV.

4.2. Particle Swarm Optimization

PSO is a population based stochastic algorithm modeled from the observation and simulation of bird flocks behavior [5]. The difference to evolutionary computation is that, in PSO, each particle benefits from its own previous solutions (historic) and there is no such approach in evolutionary methods [2]. The model of adaptive culture and particle swarms that drives PSO is based on three main principles [5]: to evaluate; b) to compare; c) to imitate.

4.3. Clustering using PSO

A PSO-based Clustering Algorithm (PSOCA) can be defined as follows [2]: in the context of data clustering, each particle represents a solution to the clustering problem and, thus, a swarm represents a set of candidate data clusterings. The quality of the particle's clustering solution can measured using validation indexes.

5. Experimental results and discussions

The Ruspini data set contains 75 patterns, distributed in four classes. The test image is a 50 x 50 pixels, RGB color image.

The encoding scheme of the two algorithms is very similar: each pattern relates to the number of the associated cluster. In GKA this is accomplished by the chromosome data structure, and in PSOCA there is the particle. Both algorithms work with the definition that every member of the population (chromosome or particle) carries a potential solution for the problem. The whole population of solutions (the population itself in GKA, and the swarm in PSOCA) works together to achieve better results for the optimization problem. On data clustering the validation indexes are MSE and Dunn's Index (DI); and on image segmentation they are MSE and Xie Beni's (XB) index, because DI is hardly time consuming.

5.1. Data Clustering results

GKA parameters are: validation index (vi), stop criterion (1 for one step k-means, 0 for 50 step maximum k-means), number of generations (gmax), number of population individuals (np), mutation probability (mp). Table 1 shows that, for each validation index, the better value of objective function is always found, despites the stop criterion. Time varies as the parameters change. It can be seen that a small number of generations, small population and small probability offers very good results.

Table 1. GKA clustering results for Ruspini dataset

n	vi	stop	k	gmax	np	mp	time(s)	f
1	mse	1	4	5	5	.025	.07	.20182
2	mse	0	4	5	5	.5	.11	.20182
3	di	1	4	5	5	.025	.36	.52101
4	di	0	4	5	5	.5	.4	.52101

Table 2. PSOCA clustering results for Ruspini dataset

n	vi	stop	k	p	time(s)	f
1	mse	1	4	5	.23	.20182
2	mse	0	4	5	.26	.20182
3	di	1	4	5	.82	.52101
4	di	0	4	5	.9	.52101

PSOCA parameters are: validation index (vi), stop criterion (1 for one step k-means, 0 for 50 step maximum k-means), number of particles (p). Table 2 shows that a small number of particles offer very good results for all validation indexes.

5.2. Image Segmentation results

Table 3 shows results for GKA and Table 4, shows results for PSOCA. It can be seen that both methods are time consuming in image segmentation and the main consideration is that, in booth algorithms, the best results are achieved by configurations (parameters) with high number of iterations, small number of population and small mutation probability (in case of GKA) and high number of particles (in case of PSOCA).

Table 3. GKA image segmentation results

n	vi	stop	k	gmax	np	mp	time(s)	f
1	mse	0	8	5	5	.025	29.79	13.774
2	mse	1	8	20	5	.025	7.3	15.214
3	xb	0	8	20	5	.025	136.44	239.771
4	xb	1	8	20	5	.025	8.76	221.893

Table 4. PSOCA image segmentation results

n	vi	stop	k	p	time(s)	f
1	mse	1	8	5	20.95	15.3466
2	mse	1	8	20	32.15	15.4035
3	xb	1	8	20	62.26	237.5309
4	xb	0	8	20	469.69	239.6779
5	di	1	8	5	4950	.0087

6. Conclusions

The present paper did show two natural computing methods for data clustering and image segmentation, one based on Genetic Algorithms and the other based on Particle Swarms. The clustering results show that both methods performed very well, finding the optimal clustering. In general, because of computational cost, GKA performed faster than PSOCA.

7. References

[1] A. K. Jain, M. N. Murty, and P. J. Flynn, *"Data clustering: a review", ACM Computer Surveys (31),* ACM Press, 1999, pp. 264-323.

[2] M. G. Omram, A. Salman, and A. P. Engelbrecht, *"Dynamic clustering using particle swarm optimization with application in image segmentation", Pattern Anal. Applic. (8)*, Springer-Verlag, London, 2006, pp. 332-344.

[3] L. N. de Castro, *"Fundamentals of natural computing: an overview", Physics of Life Reviews 4*, Elsevier, 2007, pp. 1-36.

[4] K. Krishna, and M. N. Murty, *"Genetic K-Means Algorithm", IEEE Trans. Systems, Man, and Cybernetics – Part B: Cybernetics*, vol. 29, no. 3, June 1999, pp. 433-439.

[5] J. Kennedy, R. C. Eberhart, and Y. Shi, *Swarm intelligence: collective, adaptive*, The Morgan Kaufmann Series in Evolutionary Computation, Morgan Kaufmann, New York, 2001.