

**A NEW DYNAMIC EVOLUTIONARY CLUSTERING
TECHNIQUE. APPLICATION IN DESIGNING RBF NEURAL
NETWORK TOPOLOGIES.
II. NUMERICAL EXPERIMENTS**

D. DUMITRESCU AND KÁROLY SIMON

ABSTRACT. Recently a new evolutionary optimization metaheuristics, the Genetic Chromodynamics (GC) has been proposed. Based on this metaheuristics a dynamic clustering algorithm (GCDC) is proposed. This method is used for designing RBF neural network topologies. Complexity of these networks can be reduced by clustering the training data. The GCDC technique is able to solve this problem. In Part I the GCDC technique is presented. It is described, how this method could be used for designing optimal RBF neural network topologies. In Part II some numerical experiments are presented. The proposed algorithm is compared with a static clustering technique, the generalized k -means algorithm.

Keywords and phrases: Dynamic evolutionary clustering, Genetic Chromodynamics, designing neural networks, RBF neural networks.

1. INTRODUCTION

Recently a new evolutionary search and optimization metaheuristics - called Genetic Chromodynamics (GC) (see [4, 14]) - has been proposed. Based on this theory a clustering method is proposed. This GC-based dynamic clustering technique - called GCDC - is described in [9]. The proposed algorithm can be successfully used for designing optimal RBF neural network topologies.

In this Part some numerical experiments and obtained results are presented. GCDC is used for clustering two-dimensional input data. The use of GCDC for

Received by the editors: February 18, 2005.

2000 *Mathematics Subject Classification.* 68T05, 68T20, 91C20, 92B20.

1998 *CR Categories and Descriptors.* I.2.6 [**Artificial Intelligence**]: Learning – *Connectionism and neural nets*; I.5.3. [**Pattern Recognition**]: Clustering – *Algorithms* .

designing optimal RBF neural network topologies is investigated. The method is compared with a static clustering technique, the generalized k -means algorithm [17].

In the next section the GCDC method is tested on two-dimensional input data. The behavior of the fitness function is investigated. Section 3 presents how this method can be used for designing RBF neural networks. GCDC is used for clustering training data. The topology of the RBF network is designed based on the obtained results. In the experiment presented in Section 4 the GCDC method is compared with the generalized k -means clustering algorithm .

2. EXPERIMENT 1

From the two-dimensional input space 19 data points ((x, y) pairs, where $x \in \{100, \dots, 300\}$ and $y \in \{100, \dots, 300\}$) organized in 5 clusters are considered.

GCDC is used for clustering this data set. The parameters of the method are:

- initial population size: 38;
- parameters for the fitness function: $\alpha = 2, C = 140$;
- mutation step size: $\sigma = 10$;
- merging radius: $\varepsilon = 25$.

After 45 iterations the correct number of clusters is determined by the GCDC method. The algorithm detects existing clusters and corresponding centers. The obtained results are presented in Figure 1.

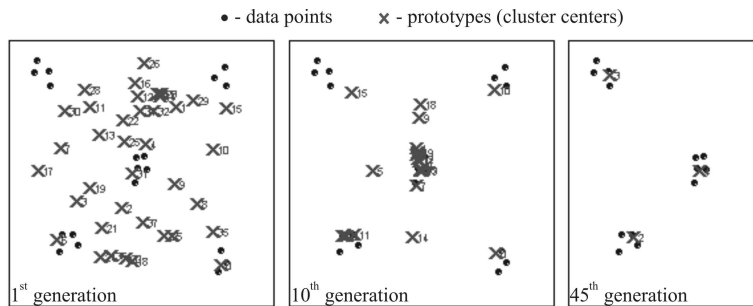


FIGURE 1. Convergence of the GCDC algorithm: two-dimensional input data, 19 data points organized in 5 clusters

More tests with different parameters for the fitness function are performed. The behavior of the fitness function is presented in Figure 2, Figure 3, Figure 4 and Figure 5.

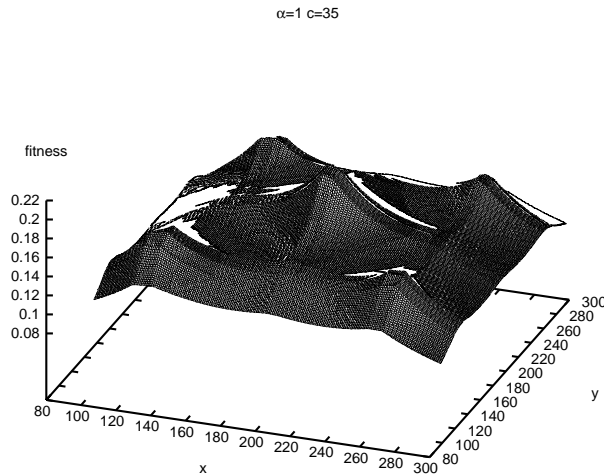


FIGURE 2. Fitness landscape for $\alpha = 1, C = 35$

3. EXPERIMENT 2

RBF neural network is used for approximating the function:

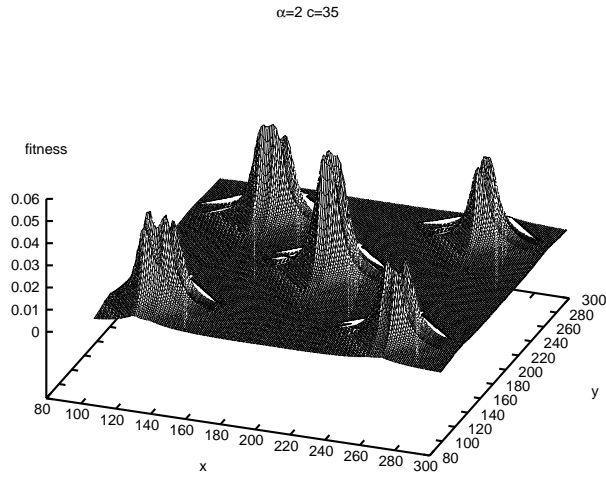
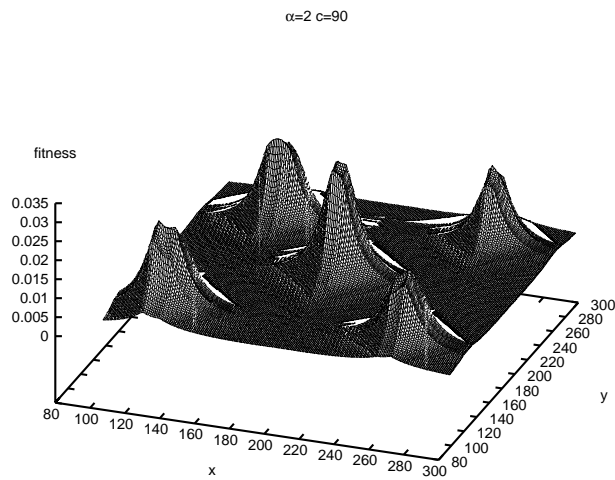
$$f : [0, 9.5] \rightarrow R, f(x) = 2 \cdot \sin \left(\ln(x) \cdot e^{\cos\left(\frac{x}{2}\right)} \right).$$

3.1. Experimental Conditions. From the interval $[0, 9.5]$ 200 points are considered as training samples. GCDC is used for clustering training data.

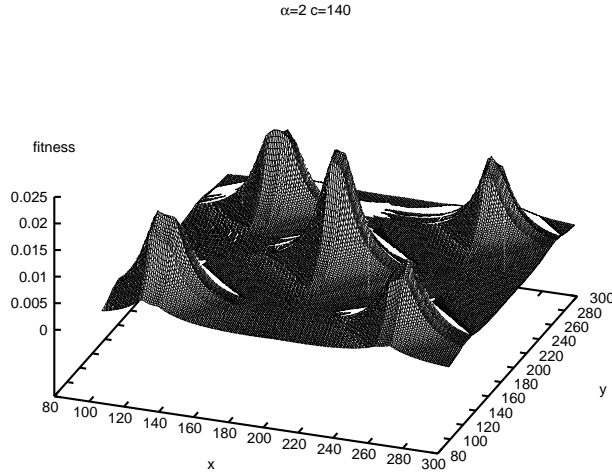
The obtained centers are used as center parameters for the RBF network. The number of processor units in the hidden layer of the network is equal with the number of centers determined by the GCDC method.

Parameters for GCDC:

- initial population size: 400;
- parameters for the fitness function: $\alpha = 1, C = 0.00001$;
- mutation step size: $\sigma = 0.0001$;
- merging radius: $\varepsilon = 0.05$.

FIGURE 3. Fitness landscape for $\alpha = 2, C = 35$ FIGURE 4. Fitness landscape for $\alpha = 2, C = 90$

Gaussian activation functions are used. The parameters for the learning algorithm are:

FIGURE 5. Fitness landscape for $\alpha = 2$, $C = 140$

- learning rate: 0.1;
- maximum number of learning epochs: 10000.

The *generalization error* is calculated using $M = 400$ inputs (that do not belong to the training set) from the interval $[0,9.5]$. The following formula is used:

$$E_g = \frac{1}{M} \sum_{i=1}^M (z_i - y_i)^2,$$

where z_i is the expected output and y_i is the network output.

3.2. RBF networks obtained by using GCDC. RBF network has been trained using 10 data sets. Each training set consists of 200 points from the interval $[0,9.5]$. In each set the points are organized in 50 well-separated clusters. For each set the GCDC method is performed and RBF neural network topologies are created based on the returned results.

In 5 cases the number of centers determined by GCDC is 50. In other 5 cases there is a little difference (maximum +4). For some classes more centers are considered. These differences have only minor effects on the network topologies. There is no situation where the number of clusters determined by GCDC is less than 50 (the optimal number of clusters).

After training the obtained RBF networks, the mean generalization error is 0.539953496. Satisfactory approximation results are obtained (Figure 6).

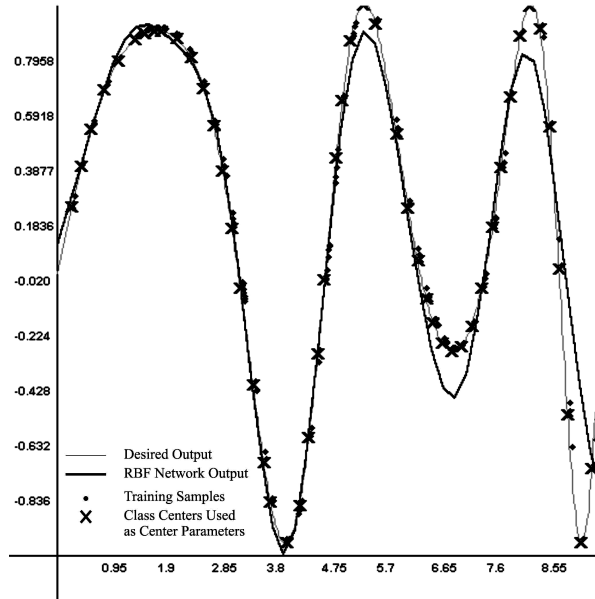


FIGURE 6. 200 training samples organized in 50 clusters, centers determined by the GCDC technique, output of the RBF network after 10000 training epochs.

3.3. RBF networks obtained by using randomly generated centers. A training set of 200 points organized in 20 clusters is considered. 20 centers are randomly selected from this set. The RBF network is designed using these centers. The procedure is repeated 10 times. After training the obtained RBF networks the mean generalization error is 0.634810589.

The GCDC technique is performed for clustering the same data set. The method finds 20 clusters and corresponding centers. Based on the returned results a RBF neural network is designed. After 10000 learning epochs the 0.591574517 generalization error is achieved. Better result is obtained using GCDC than using randomly selected centers.

4. EXPERIMENT 3

A RBF Neural Network is used for approximating the function:

$$f : [0, 1] \rightarrow R, f(x) = \left(x - \frac{1}{3}\right)^3 \cdot \frac{1}{27}.$$

The GCDC technique is compared with the generalized k -means algorithm.

4.1. Experimental Conditions. A training set consisting of 100 data points organized in 18 clusters is considered.

For k -means algorithm the number of centers is randomly generated in the range 10-25 (we assume that there are more than 10 and less than 25 clusters). 10 tests with 10 different values for the number of centers are performed.

The parameters for the GCDC algorithm are:

- initial population size: 200;
- parameters for the fitness function: $\alpha = 1, C = 0.00001$;
- mutation step size: $\sigma = 0.00001$;
- merging radius: $\varepsilon = 0.02$.

The learning rate for the training process is fixed to 0.1. The learning process will stop if the 0.00005 global learning error is achieved.

The generalization error is calculated using $M = 400$ inputs from the interval $[0, 1]$.

4.2. Obtained Results and Conclusions. The results obtained using the k -means algorithm are presented in Table 1. The mean generalization error is: 0.002228871.

GCDC detects 18 clusters and corresponding centers (Figure 7). Using these 18 centers for designing the RBF neural network the learning error of 0.00005 is achieved in 10945 epoches. The generalization error is 3.442700794496429E-4.

A better result is obtained using GCDC than using k -means. The method is able to determine the optimal number of the centers. Using the k -means method much better result is obtained by using 18 or greater value for the number of centers, than using 17 or a smaller value (18 was the real number of the centers). The learning process is thus very sensitive to the number of clusters.

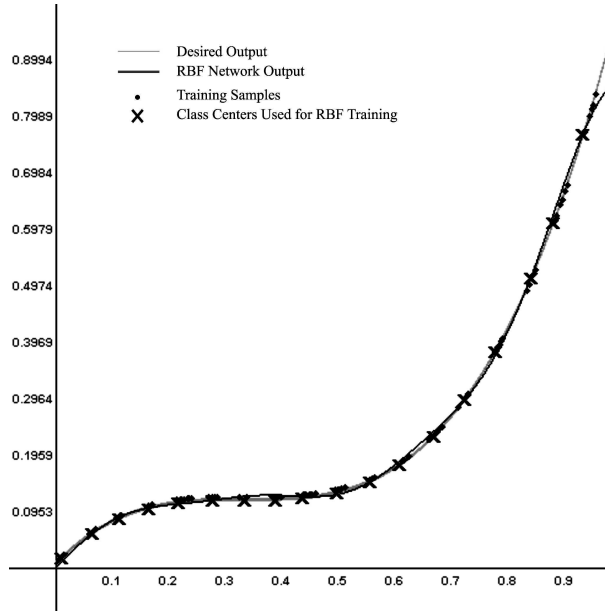


FIGURE 7. 100 training samples organized in 18 clusters, centers determined by the GCDC technique, output of the RBF network after 10945 training epochs.

5. CONCLUSIONS

Based on the GC metaheuristics, GCDC is a new evolutionary technique for dynamic clustering. Experimental results indicate that GCDC could be a powerful instrument for data clustering.

The use of GCDC for designing optimal RBF neural network topologies is investigated. Better results are obtained than using standard methods.

REFERENCES

- [1] Broomhead D. S., Lowe D.; *Multivariable Functional Interpolation and Adaptive Networks*, Complex Systems, 2 (1988), pp. 321-355.
- [2] Dumitrescu D.; *Algoritmi Genetici și Strategii Evolutive - Aplicații în Inteligența Artificială și în Domenii Conexe*, Editura Albastra, Cluj Napoca, 2000.
- [3] Dumitrescu D., Lazzarini B., Jain L. C., Dumitrescu A.; *Evolutionary Computation*, CRC Press, Boca Raton, 2000.

No. of Centers	No. of Epoches	Generalization Error
10	42386	0.003929447755582894
11	26312	0.0039125335843709025
12	15889	0.0038635588999552293
14	8218	0.0037191067346458145
16	2153	0.0028095882895919533
17	2479	0.002400189413222201
18	5466	7.485072155731134E-4
19	10208	5.057298901372404E-4
20	10017	2.240292093213028E-4
23	4918	1.76023288279397E-4

TABLE 1. Generalization errors obtained in 10 runs using the generalized k -means algorithm and 10 different values for the number of centers

- [4] Dumitrescu D.; *Genetic Chromodynamics*, Studia Univ. Babeş-Bolyai, Ser. Informatica, 35 (2000), pp. 39-50.
- [5] Dumitrescu D.; *A New Evolutionary Method and its Applications in Clustering*, Babeş-Bolyai University, Seminar on Computer Science, 2 (1998), pp. 127-134.
- [6] Dumitrescu D., Simon K.; *Evolutionary Clustering Techniques for Designing RBF Neural Networks*, Babeş-Bolyai University, Seminar on Computer Science, (2003).
- [7] Dumitrescu D., Simon K.; *Reducing Complexity of RBF Neural Networks by Dynamic Evolutionary Clustering Techniques*, Proceedings of the 11th Conference on Applied and Industrial Mathematics, (2003).
- [8] Dumitrescu D., Simon K.; *Genetic Chromodynamics for Designing RBF Neural Networks*, Proceedings of SYNASC, (2003).
- [9] Dumitrescu D., Simon K.; *A New Dynamic Evolutionary Clustering Technique. Application in Designing RBF Neural Network Topologies. I. Clustering Algorithm*, Studia Univ. Babeş-Bolyai, Ser. Informatica, (2004).
- [10] Enăchescu C.; *Caracterizarea Reţelelor Neuronale ca şi Metode de Aproximare- Interpolare*, Petru Maior University, Buletinul Stiintific, 7 (1994).
- [11] Enăchescu C.; *Elemente de Inteligenţă Artificială*, Petru Maior University, Tg. Mureş, 1997.
- [12] Haykin S.; *Neural Networks*, Macmillan College Publishing Company, New York, 1994.
- [13] Moody J., Darken C.; *Fast Learning in Networks of Locally Tuned Processing Units*, Neural Computation, 1 (1989), pp. 281-294.
- [14] Oltean M., Groşan C.; *Genetic Chromodynamics Evolving Micropopulations*, Studia Univ. Babeş-Bolyai, Ser. Informatica, (2000).

- [15] Poggio T., Girosi F.; *Networks for Approximation and Learning*, Proceedings of IEEE, 78 (1990), pp. 1481-1497.
- [16] Powell M. J. D.; *Radial Basis Functions for Multivariable Interpolation: A review*, in Algorithms for Approximation, J. C. Mason and M. G. Cox, ed., Clarendon Press, Oxford, 1987, pp. 143-167.
- [17] Schreiber T.; *A Voronoi Diagram Based Adaptive k-means Type Clustering Algorithm for Multidimensional Weighted Data*, Universitat Kaiserslautern, Technical Report, (1989).
- [18] Selim S. Z., Ismail M. A.; *k-means Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality*, IEEE Tran. Pattern Anal. Mach. Intelligence, PAMI-6, 1 (1986), pp. 81-87.
- [19] Simon K.; *OOP Pentru Calculul Neuronal*, Petru Maior University, Dipl. Thesis, 2002.
- [20] Simon K.; *Evolutionary Clustering for Designing RBF Neural Networks*, Babeş-Bolyai University, MSc. Thesis, 2003.

"BABEŞ-BOLYAI" UNIVERSITY, FACULTY OF MATHEMATICS AND COMPUTER SCIENCE, COMPUTER SCIENCE DEPARTMENT, CLUJ NAPOCA, ROMANIA

E-mail address: `ksimon@nessie.cs.ubbcluj.ro`