

Setting Initial Conditions for the RCE Model

François Blayo [†] and Michel Verleysen [‡]

[†] Swiss Federal Institute of Technology Lausanne
Microcomputing Laboratory
INF — Ecublens
CH - 1015 Lausanne

[‡] Catholic University of Louvain
Microelectronics Laboratory
3 Place du Levant
B - 1348 Louvain-la-Neuve
Currently invited professor at [†]

Abstract

The RCE algorithm, used for classification tasks, is based on incremental modification of a network structure. Despite its simplicity, it appears to be very efficient for practical applications, but requires an important number of internal units. In this paper, we propose to review the main drawbacks due to network construction method, and we propose a method to enhance the initialization phase in order to obtain a better input distribution coverage, with a reduced number of internal units.

1. INTRODUCTION - THE RCE MODEL.

The Restricted Coulomb Energy (RCE), proposed by Reilly, Cooper and Elbaum, is one of the first incremental models of neural networks. In this model, decision units are characterized by their influence region, defined by an hypersphere around the unit, whose radius is equal to the threshold of the unit. The state space is then divided into zones, each dominated by different decision units. New units are created with an initial chosen radius if a presented template does not fit into one of the influence regions of the units associated with the correct class; on the other hand, radii associated to units belonging to a wrong class but whose influence regions include the presented pattern are lowered to avoid this situation.

2. ADVANTAGES AND DRAWBACKS

The most important advantages reside in the simplicity of the algorithm and its ability to modify the number of units. The decision regions are built with respect to a threshold associated to each decision unit. The algorithm has only to decide whether a new point belongs to an existing class and, in this case, if it is correctly classified. Then, the decision of creation of a new decision unit is taken, according to the result of this comparison. A new decision unit defines a new hypersphere whose center coordinates are chosen as the actual input prototype.

The creation of class units is also decided, by a simple comparison between the activity of the class units and the desired class given by the supervisor.

The whole set of templates is presented as many times as necessary to obtain a stable configuration (no more units creation, no more radii decrease). Such a procedure guarantees the convergence. The worst case would be the creation of as many decision units as there are prototypes used for the learning. As the set of prototypes is finite, and also because the radius associated to each decision unit always decreases, the algorithm converges in a finite number of steps.

With this procedure, the ratio of correct classification (number of correctly classified templates versus the number of templates) for the learning set is 100%. The generalization ratio is however usually low, due to the following drawbacks.

First of all, the algorithm is sensitive to initial conditions. This includes the initial radius associated to each decision unit, and the choice of the order of presentation of the prototypes.

1 0 1
5 0 1

1 1 0
1 2 0

mic Time-
systems 1,

tations by
-318-365.
ve Neural

omputers

ing Fully

rtment of

ornia, San

works for
en, et al.,

Tech. Rep.

2.1 Sensitivity to initial values of the radius

Associated to each decision unit, an initial value of the radius is chosen. During the adaptation phase, the algorithm calibrates this radius in order to correctly classify the presented prototypes, with respect to the existing hyperspheres. With a large radius, each initial hypersphere covers a large part of the input space. Considering the learning set as representative of the distribution (which is a condition difficult to ensure), the providing generalization capability can be very poor since overlapping regions are large; but the number of units created during the learning phase can be optimized. The generalization capability is defined as the potential ability of the network to correctly classify vectors never presented during the learning phase. On the opposite, a small initial radius in the same conditions decreases the overlapping regions but enhances the number of decision units. A pathological example is given in figure 1 : the three templates $(-1,0)$, $(1,0)$ and $(0,1)$ belong to class one, while the fourth template $(0,-1)$ belongs to class two. Initial radii are set to 2 in figure 1 a), and to 1 in figure 1 b).

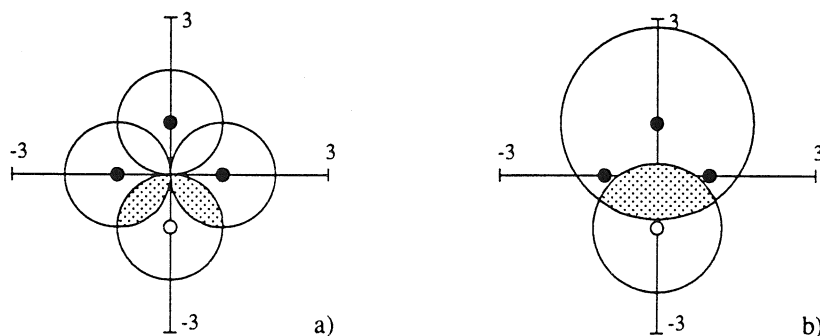


Fig. 1 Sensitivity to the initial radius

With the standard algorithm, the initial radius is randomly chosen, giving sometimes a very poor coverage of the input space. An important research effort would be the study of the initial radius, given by an adequate preprocessing, for instance, or by the evaluation of a criterion for each input vector.

2.2 Dependency on the presentation order of vectors in the training set

The presentation order of the training set vectors has also an important consequence on the space coverage. Indeed, each weight vector, associated to each decision unit, is fixed to the input vector value which lead to the creation of the unit. A chosen reference vector thus never changes. Figure 2 shows this dependency, with the same training set as in figure 1 (initial radii are set to 2), but presented in the order $(0,1)$, $(0,-1)$, $(1,0)$ and $(-1,0)$ for figure 1 a), and in the order $(1,0)$, $(0,-1)$, $(0,1)$ and $(-1,0)$ for figure 1 b).

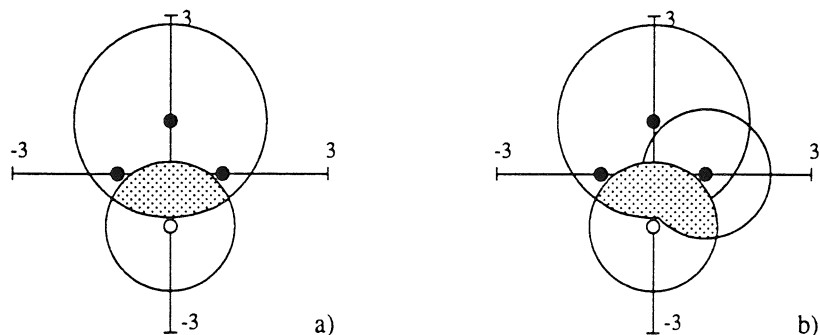


Fig. 2 Sensitivity to the order of presentation

n phase, the
spect to the
nput space.
to ensure),
ge; but the
apability is
during the
overlapping
gure 1 : the
belongs to

3
1

b)

very poor
adius, given
ctor.

n the space
ector value
shows this
in the order
gure 1 b).

3
1

b)

In this paper, we propose a simple procedure to choose the initial values of the decision unit vectors. This procedure reduces the number of decision units created during the learning phase, and increases the convergence of the algorithm. It also decreases the size of the overlapping regions. This procedure will be described in section 3.

2.3 Regions without covering.

Another problem comes from the symmetry of the influence region associated to each decision unit. The input space is not completely covered by the superposition of hyperspheres associated to each weight vector of the decision units. With a Voronoi tessellation, such a coverage should be better, but involves the knowledge of the neighbour of each reference point. In this last case, the procedure is not local and therefore not incremental. Influence regions formed with circles and with a Voronoi tessellation are respectively depicted in figure 3 a) and b). This problem has been addressed by (Alpaydin, 1990), and a first solution has been proposed with the Grow And Represent (GAR) algorithm.

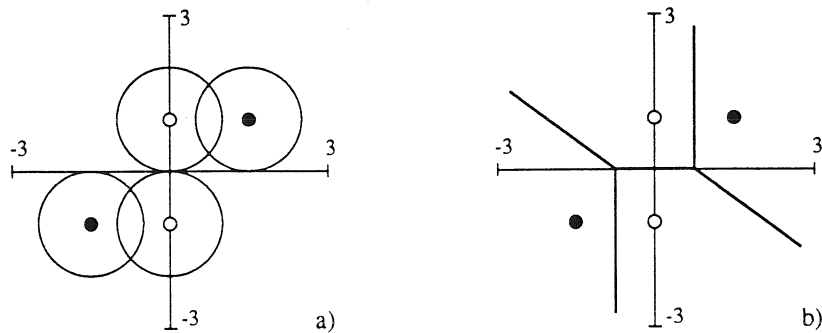


Fig. 3 Circular influence regions and Voronoi tessellation

3. A SIMPLE PROCEDURE FOR INITIAL CONDITIONS

In order to enhance the RCE algorithm performances, we propose a simple procedure which can be considered as a *preprocessing* performing some estimation of the centroid of sub-distributions, each one being defined as the points surrounded by the hypersphere created by an initial reference vector.

This preprocessing has the possibility to "move" the reference vector, in order to increase the a priori intra-class probability (maximize the correct classifications), and to reduce the a priori extra-class probability (minimize the misclassifications). This procedure is described in the following.

The preprocessing is splitted in two phases: an initial distribution of the reference vectors defining some dominance hyperspheres and an adaptation of these reference vectors according to the points surrounded by the hyperspheres.

3.1 Choice of the initial distribution of reference vectors.

The initial distribution of the reference vectors is similar to the RCE procedure, with only one difference. When an input vector is wrongly classified in a dominance hypersphere, its radius is not reduced, but a new dominance hypersphere with the same initial radius is created; as in the RCE algorithm, its center is initially set to the input vector.

3.2 Adaptation of the reference vectors

During the second phase, the reference vectors are moved, according to the following equations:

- for each presented input vector x , if it belongs to an hypersphere which is part of the correct class, move the weight vector of the hypersphere in the direction of x , with a gain parameter α_c .

$$W_i(t+1) = W_i(t) + \alpha_c(x - W_i(t)) \quad (\text{Eq. 1})$$

$$0 \leq \alpha_c \leq 1$$

- if the input vector x belongs to an hypersphere which is part of a wrong class, move away the weight vector in the opposite direction from the vector x , with a gain parameter α_a .

$$W_i(t+1) = W_i(t) - \alpha_a(x - W_i(t)) \quad (\text{Eq. 2})$$

$$0 \leq \alpha_a \leq 1$$

This procedure is very similar to the LVQ adaptation rule (Kohonen, 1988), with a different gain parameter for each equation. The adaptation procedure is repeated for a defined number of iterations (*epoch* parameter); it must be chosen at the beginning of the pre-processing. A better adaptation occurs with small values of the gain parameters (typically 0.002...0.01), and with a large number of epochs (10...50).

After the preprocessing phase, the original RCE algorithm can be used to classify the entire training set, i.e to define the decision region associated to each class, and to add decision units if necessary. Two examples are given in section 4, in order to compare the results obtained with the RCE with and without the proposed preprocessing.

4. SIMULATION RESULTS

Two examples are given in the following. They are intended to give a qualitative measure of the RCE improvement given by the proposed method.

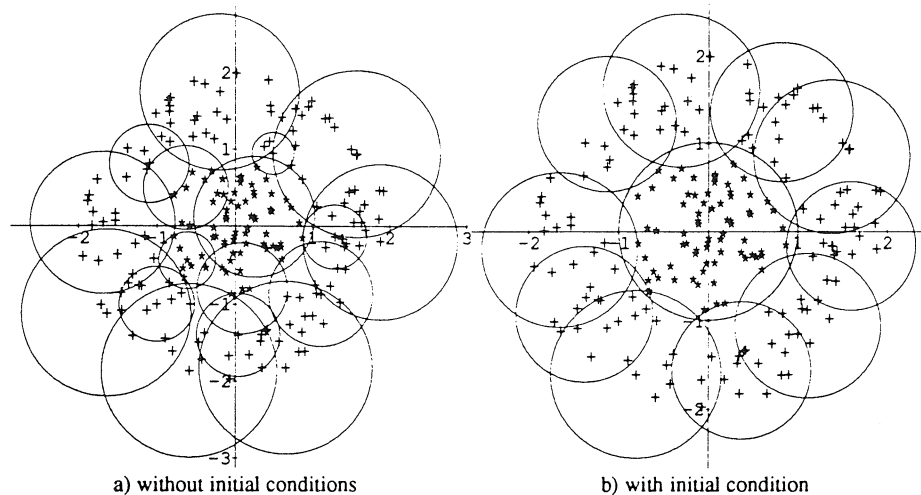


Fig. 4 Example on circular and ring distribution

The example of figure 4 shows a circular distribution, which is a very simple problem for the RCE algorithm, and a ring distribution, which is more difficult to cover. The RCE algorithm, without preprocessing, has allocated 4 units to cover the circular distribution, and 13 units for the ring. The variance of the circle radii is important. With preprocessing, the circular distribution is covered with only one unit, and the other 10 units have almost identical radii. In that sense, the covering is much better.

ss, move

b)

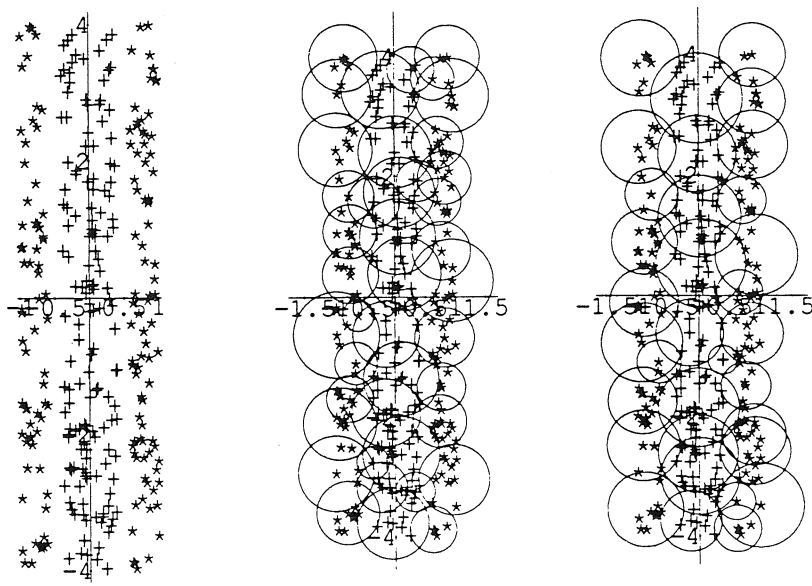
weight

c)

ent gain
s (epoch
urs with
10...50).

ning set,
ry. Two
without

the RCE



a) distribution of classes b) result without initial conditions c) result with initial conditions

Fig. 5 Example of strip distribution

In this last example, one can see that the variance of the radii are different between the 5 b) and 5 c) figures. The number of units is also different: 42 units for the 5 b) example, compared to the 33 used for the 5 c). Again, the coverage with preprocessing seems better than the one obtained with the original RCE algorithm.

5. CONCLUSION AND FUTURE WORK

The RCE algorithm is an algorithm of the "Region of Influence Algorithms" class. It can be successfully used for classification purpose, but suffers from some problems due to the limited knowledge about the real distribution of points during learning. This fact is especially important during the initialization phase. In order to overcome a part of these limitations, we propose a method to improve the initial position of the hyperspheres centers. It is based on a LVQ-like adaptation rule, used to move the centers of the hyperspheres dominating a sub-distribution. With such a method, we observe a reduction of the created units, and a better coverage of the distribution.

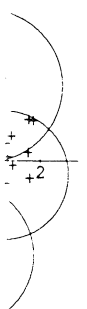
Despite its simplicity, this method is a first step to improve the performances of an incremental learning algorithm. Such an approach must be completed with other studies to deal with the other drawbacks underlined in this paper. Further work may concern the choice of the initial radii, and criteria to avoid that several units include the same regions during the first initialization phase. Part of this future work will be achieved in the ESPRIT3-BRA ILENA-Nerves2 project.

6. REFERENCES

Alpaydin, E. (1990), Neural models of incremental supervised and unsupervised learning, PhD Thesis No. 863, Swiss Federal Institute of Technology, Lausanne.

Kohonen, T., (1988), Self-organization and associative memory, 2nd Edition, Springer Verlag, Berlin

Reilly, D. Cooper, L., Elbaum, C., (1982), A neural model for category learning, *Biol. Cyb.*, **45**, N1, 35-41



the RCE
without
ing. The
with only
etter.