An Intuitionistic Fuzzy RBF Network

Ludmila Kuncheva*
Department of Electrical & Electronic Engineering
Imperial College, Exhibition Road, London SW7 2BT, UK
e-mail: L.Kuncheva@ic.ac.uk

Krassimir Atanassov
CLBME, Bulgarian Academy of Sciences
Acad. G. Bonchev Street, Block 105, 1113 Sofia, BULGARIA
tel. +359 2 713-3602, fax. +359 2 72 37 87
e-mail: krat@bgcict.acad.bg

ABSTRACT: A radial-basis-function (RBF) network is considered with activation functions taking highly nonsymmetric form, specific for each kernel, thereby making the attraction regions varying in shape and size. The representation of the function is inspired by the intuitionistic fuzzy set theory which is used to account both for activation and restraining potentials at a node. We believe that by using fewer hidden nodes in such a configuration, we can approximate complex classification boundaries. The training algorithm used here is a version of simulated annealing. The experimental results with the two-spirals data set demonstrate the increased capability of the proposed RBF network.

1 INTRODUCTION

Radial Basis Function networks have been shown to surpass other network architectures with respect both to classification accuracy and training ease. They have been proved to be asymptotically Bayes-optimal, being "relatives" to Parzen's density function estimators [Coultrip 1994].

In general, RBF network use a radially symmetric kernel function $K_j((x - c_j)/\sigma)$ at the hidden layer. Usually, the closer x to c_j, the higher the activation, whatever distance is used. Some studies show that RBF networks have the universal arrpoximation capabilities. It has been shown elsewhere that many kernel function models satisfy a set of conditions making RBF networks universal approximators property. In particular, any function $K(z)$, $z \in \mathcal{R}^n$ can be used, provided it is integrable bounded, such that it is continuous almost everywhere and its integral over the feature space $\mathcal{R}^n$ is nonzero. The kernel functions used at the hidden nodes are of the same type and with the same value of the smoothing parameter but the result can be extended (and sharpened, [Park 1993]) to functions with varying σ from node to node. It turns out that the radial symmetry is not necessary for the proof.

Various RBF shapes have been considered in the literature: general exponential, Gaussian, elliptical [Tattersal 1994], fuzzy [Kim 1995], cubic [Halgamuge 1995], B-spline [Brown 1995], etc. In these works the same shape of the RBFs for all hidden units is implied with only the parameters modified. Here we suggest the use of different shapes, *specific* for each hidden node. The representation of the function is inspired by the intuitionistic fuzzy set (IFS) theory which is used to account both for activation and restraining potentials at a node. We believe that by using fewer hidden nodes in such a configuration, we can approximate complex classification boundaries.

*On leave from CLBME, Bulgarian Academy of Sciences

The network structure and functioning is described in Section 2. Some experimental results are presented in Section 3.

2 THE RBF NETWORK

Let $x \in \mathcal{R}^n$ be a feature vector, and $\Omega = \{\omega_1, ..., \omega_M\}$ be a set of classes defined over $\mathcal{R}^n$. The proposed IFS RBF network is sparsely connected, i.e. each node at the hidden layer is connected with only one output node. Each such node $N_j, j = 1, ..., L$ with a corresponding center vector $c_j \in \Re^n$ is supplied with two kernel functions: $\mu(x, c_j)$ for activation, and $\nu(x, c_j)$ for restraining, respectively, $\mu(x, c_j), \nu(x, c_j) \in [0, 1]$. In the IFS theory there is a requirement that $\mu(x, c_j) + \nu(x, c_j) \leq 1$, which corresponds to the logical nature of the two degrees. Here we have overlooked this because we interpret μ and ν only as components of the transition function and not in their logical context. A neuron at the hidden layer renders certain degree of "support" to the newcomer x for the class ω_i, which the neuron is responsible for. This degree is formulated on the basis of $\mu(x, c_j)$ and $\nu(x, c_j)$. Output nodes simply sum up the support given to the discussed object. The object x is labeled as belonging to the class with the maximal support.

The output of the hidden-layer node N_j is

$$y_j^h(x) = F(\mu(x, c_j), \nu(x, c_j)), \quad j = 1, \ldots, L, \tag{1}$$

where $F(.,.)$ is a function aggregating the activating and inhibitory potentials of the neuron. Note that the aggregation can be asymmetric which corresponds to the known behavior of the biological neuron.

The output nodes perform the summation

$$y_k^o(x) = \sum_{j=1}^{L} Ind(k, j)\, y_j^h(x), \quad k = 1, \ldots, M. \tag{2}$$

where $Ind(k, j)$ is an indicator function that yields 1 if node N_j corresponds to class ω_k, and 0, otherwise. Obviously, $\mu(x, c_j)$ and $\nu(x, c_j)$ can be formulated in many ways. In this study the following two functions are used

$$\mu(x, c_j) = exp(- \| x - c_j \|_{L_2}^2), \tag{3}$$

where $\| x - c_j \|_{L_2}$ is the Euclidean norm in $\Re^n$, and

$$\nu(x, c_j) = 1 - exp(- \| x - c_j \|_{L_1}^2), \tag{4}$$

with $\| x - c_j \|_{L_1}$ - the L_1 norm. In the proposed network the location of the seed points (centers of the neurons at the hidden layer) is not changed. In order to make the two functions trainable we use a weighted L_p norm:

$$\| x - c_j \|_{L_p} = \{\sum_{k=1}^{n} w_{k,j} |x^k - c_j^k|^p\}^{1/p} \tag{5}$$

where the tunable parameters are $w_{k,j}$.

The aggregation function adopted here is:

$$F(\mu(x, c_j), \nu(x, c_j)) = \begin{cases} \mu(x, c_j), & \text{for } \nu(x, c_j) \leq K_\nu \\ \mu(x, c_j)(1 - \nu(x, c_j)), & \text{otherwise} \end{cases} \tag{6}$$

where K_ν is a parameter. By this means we can obtain radial basis functions of quite exotic shapes. The rationale behind the above equation is that when the inhibitory rate is greater than a certain threshold the degree of support decreases proportionally.

Obviously, when increasing the number of trainable parameters, it can be expected that the classification ability of the scheme will grow. On the other hand, it is not clear whether the particular way of using the parameters can lead to good performance. The ability of the scheme to train is demonstrated in the next section. As the training algorithm we used simulated annealing.

3 EXPERIMENTAL RESULTS

The ability of the IFS RBF to train are illustrated using the two-spirals data set, one of the artificial benchmark data sets for neural network classifiers. As it can be seen in Fig. 1 the set consists of two intertwined spirals representing two 2-D classes.

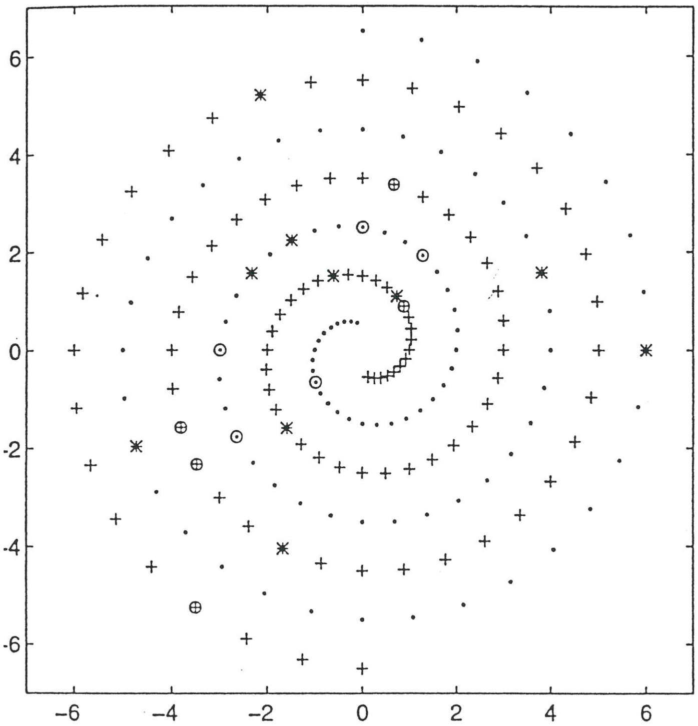

Figure 1: The two-spirals data set. The first and the second sets of prototypes are depicted by '*' and 'o', respectively

The original data set consists of two subsets: the training and test one with a negligible difference between them. Therefore the generalization ability of the proposed model cannot be properly assessed on this benchmark problem. The challenge in this problem is the classification boundary which is is highly nonlinear and intricate. The string-like structure of the classes suggests that the nearest neighbor technique ($1 - NN$) will be the perfect choice of a classifier. Indeed, both on the training set (with leave-one-out) and on the test set, $1 - NN$ leads to 100 % correct recognition. If increasing the number of neighbors ($k - NN$ rule) the performance deteriorates. It follows that the more prototypes we retain, the higher accuracy we can get. We emphasize here that the main objective has been to compare the classifiers rather than to reach the highest possible accuracy. Two sets of centers of cardinality 10 have been selected at random (called *Set 1* and *Set 2*, respectively). Three RBF configurations have been compared:

(A.) RBF with activation function of the form:

$$y_j^h(x) = exp(- \| x - c_j \|_{L_2}) \;\; j = 1, \ldots, L, \tag{7}$$

Table 1: Means and standard deviations (in %) with the 2-spirals data set (Prototype *Set 1*)

Cooling scheme	A		B		C	
	Training	Test	Training	Test	Training	Test
1	55.08 ±0.8643	55.12 ±1.1670	58.86 ±1.7473	58.98 ±2.4874	58.44 ±5.1714	58.44 ±4.6560
2	57.02 ±0.5848	55.74 ±1.6682	59.80 ±1.3620	59.70 ±1.6748	60.90 ±2.6748	60.10 ±2.3206

(B.) RBF with activation function of the form:

$$y_j^h(x) = exp(- \| x - c_j \|_{L_p}) \; j = 1, \ldots, L, \tag{8}$$

with tunable p, varying in the interval [0.5,3.5], and

(C.) The proposed IFS RBF model.

With each of these, 5 runs of the SA training algorithm, starting from different initialization, have been performed.

Tables 1 and 2 contain the means and the standard deviations of the classification accuracy on the training and on the test set with *Set 1* and *Set 2*, respectively. Two cooling schedules of the simulated annealing (SA) algorithm have been applied. In the second schedule twice more iterations at a temperature level have been performed. Figure 2 shows the classification accuracy of the best runs with configurations A, B, and C.

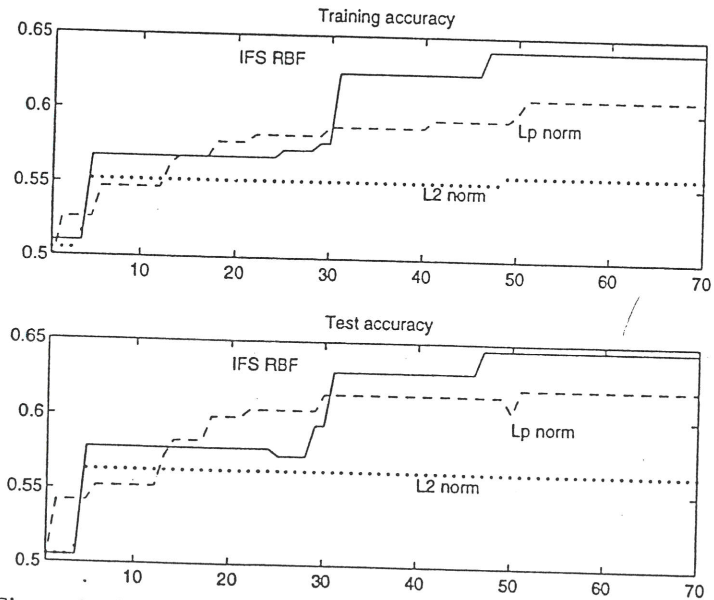

Figure 2: Classification accuracy with the 2-spirals data set (the best run with each configuration)

As expected, the results differ significantly for the two prototype sets. Since the choice of those sets has been made completely at random, the results are not supposed to indicate the classification accuracy achievable with the prototype sets of the considered cardinality.

It can be seen that the IFS RBF network can train better than its competitors. The variance of the estimate of the classification accuracy is smaller with configuration A than that with B, or C. This means

Table 2: Means and standard deviations (in %) with the 2-spirals data set (Prototype *Set 2*)

Cooling scheme	A		B		C	
	Training	Test	Training	Test	Training	Test
1	60.10 ±1.3248	60.84 ±2.3597	60.72 ±2.2510	61.04 ±1.6103	61.56 ±2.5967	61.34 ±2.4946
2	60.12 ±1.2911	60.62 ±1.7598	61.68 ±2.2174	62.52 ±3.0842	62.70 ±3.2962	62.40 ±3.3489

that, perhaps, the training scheme lets the network A achieve the best result up to its capacity. Indeed, with the second cooling scheme (which assures for a more extensive training), the results with the IFS RBF network improve while the ones with the network A are virtually the same, with both prototype sets. This is an expected result, given that more parameters are tuned with networks B and C. Besides, this improvement indicates that with a judicious selection of the cooling scheme better results can be achieved.

CONCLUSIONS

In this paper we propose an RBF network configuration with tunable function shape based on IFS representation. We illustrate the ability of the IFS RBF network to train by comparing it with two other RBF network structures on the two-spirals data set. The same training scheme based on simulated annealing algorithm has been used. The results show that the way we have incorporated the extra parameters into the radial-basis function gives the network additional flexibility and increased capacity.

ACKNOWLEDGEMENTS

The scholarship kindly provided by the Royal Society and Foreign & Commonwealth Office is greatly appreciated.

REFERENCES

Brown, M.; Bossley, K.M.; Mills, D.J.; Harris, C.J. 1995. High Dimensional Neurofuzzy Systems: Overcoming the Curse of Dimensionality. FUZZ/IEEE'95, Yokohama, Japan, pp. 2139-2146.

Coultrip, R.L.; Granger, R.H. 1994. Sparse Random Networks with LTP Learning Rules Approximate Bayes Classifiers via Parzen's Method. Neural Networks Vol. 7, pp. 463-476.

Halgamuge, S.; Poechmueller, W.; Glesner, M. 1995. An Alternative Approach for Generation of Membership Functions and Fuzzy Rules Based on Radial and Cubic Basis Function Networks. International Journal of Approximate Reasoning Vol. 12, pp. 279-298.

Kim, H.M.; Mendel, J.M. 1995. Fuzzy Basis Functions: Comparison with Other Basis Functions. IEEE Transactions on Fuzzy Systems Vol. 3, pp. 158-168.

Park, J.; Sandberg, I.W. 1993. Approximation and Radial-Basis-Function Networks. Neural Computation Vol. 5, pp. 305-316.

Tattersall, G.D.; Yi, K. 1994. Packed Hyper-Ellipsoid Classifiers. Electronic Letters Vol 30, pp. 427-428.