A CONSTRUCTIVE ALGORITHM FOR FUZZY NEURAL NETWORKS

F.M. Frattale Mascioli, G. Martinelli, and A. Rizzi

INFO-COM Dpt., University of Rome "La Sapienza"
Via Eudossiana 18, 00184 Roma, Italy
mascioli@infocom.ing.uniroma1.it

ABSTRACT

We propose a constructive method, inspired by Simpson's Min-Max technique, for obtaining fuzzy neural networks. It adopts a cost function depending on a unique net parameter. This feature allows us to apply a simple unimodal search for determining this parameter and hence the architecture of the optimal net. The algorithm shows a good behavior with respect to other methods when applied to real classification problems. Due to the adopted fuzzy membership functions, it is particularly indicated when the classes are extremely overlapped (for instance, in the case of biological data). Some results at this regard are reported in the paper.

1. INTRODUCTION

In order to improve the efficiency of neural networks, the synergism between them and other important topics has been recently exploited in the technical literature. The topics involved in this effort are based on very general principles related to both optimization and human behavior imitation. They are the *Learning Theory* and the *Fuzzy Inference Systems* [1–3]. The present work regards the development of a constructive algorithm for determining fuzzy neural networks. As suggested by Learning Theory, this approach improves the generalization capability of the net. We obtain it in connection with Simpson's Min-Max technique [4], which adopts simple fuzzy membership functions and assures low computational cost, robustness and ability to classify real data. Our algorithm minimizes a cost function depending on a unique net parameter. This feature allows us to apply a simple unimodal search for determining this parameter and hence the architecture of the optimal net (in terms of structural complexity). Simulations are described to show the good behavior of the algorithm, especially in classification problems where the classes are particularly overlapped.

2. MIN-MAX TECHNIQUE

The fuzzy neural network proposed in [4] for classification problems is characterized by its ability to learn on-line and in a single pass through the data. The net operation consists in a suitable mechanism of partitioning the input space. The regions corresponding to the classes are covered by *hyperboxes* parallel to the coordinate axes, with appropriate membership functions. The location of each hyperbox is completely defined by two extreme vertices: the 'min' and 'max' vertices. Min-Max training algorithm consists in determining the hyperboxes necessary to cover the classes. It is characterized by a three step process:

1) *Expansion:* in this step the hyperboxes already constructed are expanded in order to accommodate a new example of the training set. The expansion is limited by a maximum dimension, represented by a parameter θ to be chosen carefully ($\theta \in (0,1]$);

2) *Overlap test:* determination of overlaps among hyperboxes of different classes;

3) *Contraction:* elimination of overlaps, if they exist.

The structural complexity of the resulting neural net is directly defined by the final number of hyberboxes. In fact, the net is constituted by a hidden layer containing a neuron per hyperbox and an output layer with a neuron per class. The original Simpson's technique has been modified in relation to the problems at hand. In particular, to obtain smoother classifications *hyperellipsoids* have been adopted as membership functions in problems whith extremely overlapped classes. More precisely, the original hyperboxes are replaced by hyperellipsoids only at the end of the constructive process, in order to save the low computational cost of Simpson's technique. The general expression of the new membership functions is the following:

$$\mu(\mathbf{x}) = \frac{1}{1 + \sum_{\mathbf{i}=1}^{\mathbf{n}} \frac{(\mathbf{x_i} - \mathbf{c_i})^{2\mathbf{b_i}}}{\mathbf{a}_\mathbf{i}^2}} \quad (1)$$

Where: $\mathbf{x}_i$ is the i-th component of the input vector $\mathbf{x}$; $\mathbf{c}_i$ is the i-th component of the hyperbox center $\mathbf{c}$; the parameters $\mathbf{a}_i$, $\mathbf{b}_i$ control the shape of function $\mu(\mathbf{x})$ and are proportional to the hyperbox sizes.

3. THE CONSTRUCTIVE STRATEGY

Learning theory suggests as the objective function to be minimized during training the sum of two terms, E_S and E_C, which "measure" respectively the *error* on the training set and the *structural complexity* of the net. We note that both the complexity of the net and the accuracy of the classification of the training set depend on the maximum size allowed to the hyperboxes, which is proportional to the parameter θ. The two quantities E_C and E_S can be reasonably measured by the number N of hyperboxes (neurons of the hidden layer) and by the number M of misclassifications occurred during training. The resulting objective function is therefore:

$$F(\theta) = (1-\lambda)\cdot M + \lambda\cdot N \quad (2)$$

where λ ($\lambda \in [0,1]$) is a weight to be determined taking into account that small values of λ yield nets more complex but less erroneous (the opposite situation characterizes large values of λ). The dependence of F on θ is not smooth as a consequence of the several possible choices the Min-Max algorithm can undertake during training. Preliminary tests have in fact shown that functions $E_S(\theta)$ and $E_C(\theta)$ are affected by several local minima. Because of this irregular behavior, in order to find the optimal value θ_{min} without being trapped in local minima of F, a uniform sampling of the whole interval of interest of parameter θ is necessary. The proposed algorithm carries out this sampling in successive intervals of decreasing length and with a growing resolution. In each interval the value of θ which minimizes F is determined by considering, for all the sampled values of θ, the corresponding net and function F. The successive interval is then centered in correspondence to this minimum and sampled with a larger resolution. The number of values of θ to be considered for achieving the optimal value θ_{min} is drastically reduced by taking account that:

1) by decreasing the value of θ step-by-step, a value $\overline{\theta}$ exists such that $E_S(\overline{\theta})=0$, i.e. a "zero-errors" situation is achieved. When the value $\overline{\theta}$ is obtained, no further decrease of θ is necessary, since $F(\theta)$ could only increase with respect to $F(\overline{\theta})$. In fact: $E_S(\tilde{\theta})=0$ and $E_C(\tilde{\theta}) \geq E_C(\overline{\theta})$ for $\tilde{\theta} < \overline{\theta}$;

2) the maximum resolution to be used is chosen on the basis of the accuracy required by the training set in the input space.

4. SIMULATION RESULTS

To test the efficiency of the proposed algorithm we have carried out several experiments, some of them are described in the following:

Twin-spiral: This is a classical benchmark problem used for its extreme nonlinearity [5]. The training set consists of 194 points (X-Y real values) arranged in two interlacing spirals. The more significant results of the test are reported in Tab. 1 and in Fig. 1. They compare favorably with respect to other methods both in terms of complexity and accuracy [6].

Iris: This is a well-known taxonomic problem, often used as a benchmark. The training set consists of 150 four-dimensional patterns which represent multiple measurements of three different species of plants: *Iris setosa*, *I. versicolor*, *I. virginica* [7]. The results obtained in this case, summarized in Tab. 2, are superior to those available in the technical literature which we are aware of [8].

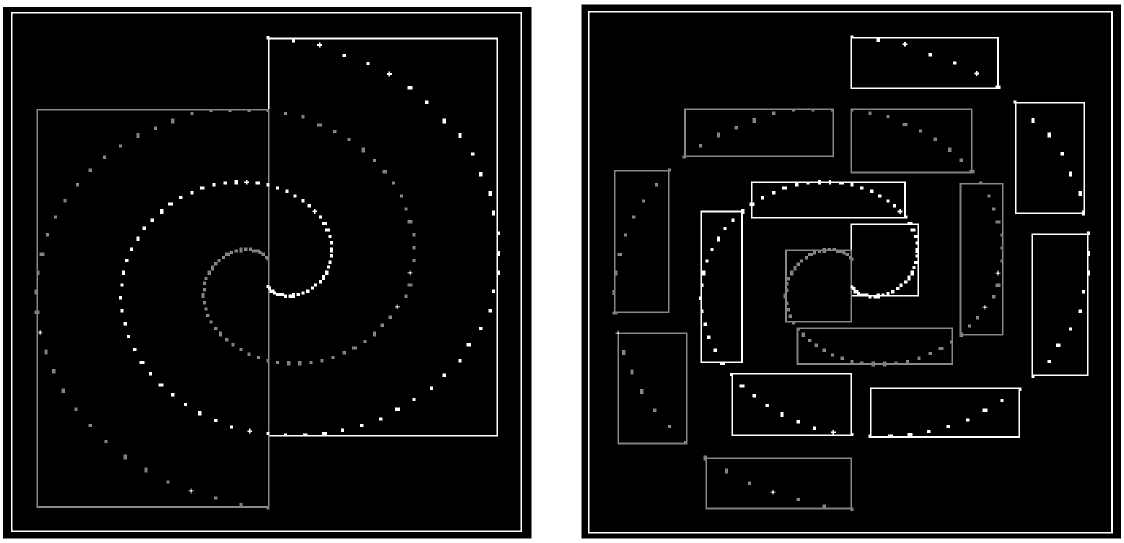
Bear skulls: Problem data consist in a set of skull measurements (basal length and zygomatic width) of two species of upper Pleistocene bears (*Ursus spelaeus*, *U. arctos*) and two present sub-species (*U. arctos alpinus*, *U. arctos marsicanus*). The best result is obtained with λ=0.5 and θ_{min}=0.43, which corresponds to a net constituted by only 6 hidden neurons. The misclassification amounts to 6.3%. This result is considered by specialists in Paleontology very interesting due to the extreme overlapping of the four classes [9].

Phoneme classification: The use of our algorithm in a hybrid ANN/HMM (Artificial Neural Network/Hidden Markov Model) system for continuous speech recognition [10, 11] has been investigated. The phoneme classifier, which constitutes the first part of the recognition system, has been trained by means of TIMIT database. Preliminary results show the superiority of our approach in terms of generalization capability, time required for training and representation of widely spreaded data, in comparison to the same system adopting a statistical classifier based on gaussian mixture.

TWIN-SPIRAL PROBLEM: CLASSIFICATION RESULTS

λ	θ_{min}	*Misclassifications*	*Rejects*	*Hidden neurons*
0÷0.1	0.19	0 %	1 %	18
0.5÷0.75	0.21	0 %	1 %	16
0.78	0.25	3.1 %	0 %	14
0.8÷0.82	0.74	26.8 %	0 %	4
0.9÷1	0.84	33.5 %	0 %	2

Table 1: The table describes the classification results obtained by the proposed constructive algorithm for the twin-spiral problem. The optimal result in terms of generalization capability is obtained for λ=0.5 and θ_{min}=0.21.



(a) (b)

Figure 1: Twin-spiral problem: (a) the distribution of the hyperboxes for λ=1 and θ_{min}=0.84; (b) the same one for λ=0.5 and θ_{min}=0.21.

IRIS DATA PROBLEM: CLASSIFICATION RESULTS

λ	θ_{min}	*Misclassifications*	*Rejects*	*Hidden neurons*
0÷0.1	0.18	0 %	2 %	23
0.15÷0.33	0.42	0.67 %	0 %	6
0.5	0.54	1.34 %	0.67 %	4
0.67÷1	1	4 %	0 %	3

Table 2: The table describes the classification results obtained by the proposed constructive algorithm for the iris data problem. The optimal result in terms of generalization capability is obtained for λ=0.15 and θ_{min}=0.42.

5. CONCLUSIONS

The proposed method yields very satisfactory results in terms of generalization, net complexity and processing time. It operates well in real classification problems; especially with biological data, where the classes are overlapped and a sharp boundary among them does not exist. The algorithm performances depend on the order of presentation of the examples of the training set. However, this factor (that requires a proper data ordering) is relevant only in the case of off-line learning. There are other factors which also affect the behavior of the algorithm, in particular the type of membership function. In fact, in several applications where the classes are particularly overlapped, the use of membership functions different from the original hyperbox, as for instance the hyperellipsoid, gives better results. Finally, we remark that constructive approach in connection with neuro-fuzzy modelling is a very important and active research topic. The Min-Max technique is particularly suited for this purpose. In fact, the hyperboxes found by the algorithm can be easily used for deriving the premise part for each *if-then* rules of the underlying model [12]. Some preliminary results at this regard are encouraging.

ACKNOWLEDGMENT

This work was supported in part by the Italian Ministry for University and Research (MURST) and the Italian Research Council (CNR).

REFERENCES

[1] D.R. Hush and B.G. Horne, “Progress in supervised neural networks”, Signal Proc. Magazine, pp. 8-39, 1993.

[2] S. Haykin, “Neural Networks: a comprehensive foundation”, Macmillan, 1994.

[3] IEEE Trans. on Neural Networks, Special issue on Fuzzy Logic and Neural Networks, 1992.

[4] P.K. Simpson, “Fuzzy min-max neural networks - Part 1: classification”, IEEE Trans. Neural Networks, Vol. 3, pp. 776-786, 1992.

[5] K.J. Lang and M.J. Witbrock, “Learning to tell two spirals apart”, Proc. of the Connectionist Models Summer School, Morgan Kaufmann, 1988.

[6] F.M. Frattale Mascioli, G. Martinelli, and G. Lazzaro, “Comparison of constructive algorithms for neural networks”, Proc. of the ICANN ‘94, Springer-Verlag, pp. 731-734, 1994.

[7] R.A. Fisher, “The use of multiple measurements in taxonomic problems”, Annals of eugenics, Vol. 7, pp. 179-188, 1936.

[8] H. Ishibuchi, K. Nozaki, and R. Weber, “Approximate pattern classification with fuzzy boundary”, Proc. of the 1993 IJCNN, pp. 693-696, 1993.

[9] L. Capasso Barbato, E. Cerilli, and C. Petronio, “Differenze morfologiche e morfometriche nei crani di Ursus spelaeus e U. arctos (Morphological and morphometric differences in skulls of U. spelaeus and U. arctos)”, Il Quaternario, Vol. 6(1), pp. 67-76, 1993.

[10] N. Morgan and H. Bourlard, “Neural network for statistical recognition of continuous speech”, Proc. of IEEE, Vol. 83, n. 5, pp. 741-770, 1995.

[11] N. Morgan and H. Bourlard, “Continuous speech recognition”, IEEE Signal Processing Magazine, pp. 25-42, May 1995.

[12] J.-S. R. Jang and C.-T. Sun, “Neuro-fuzzy modelling and control”, Proc. of the IEEE, Vol. 83, pp. 378-406, 1995.