ICANN '94

Proceedings of the International Conference
on Artificial Neural Networks
Sorrento, Italy
26–29 May 1994

Volume 1, Parts 1 and 2

Edited by
Maria Marinaro and Pietro G. Morasso

Springer-Verlag
London Berlin Heidelberg New York
Paris Tokyo Hong Kong
Barcelona Budapest

Maria Marinaro
University of Salerno
Dipartimento di Fisica Teorica e S.M.S.A
Via S. Allende
84081 Barnossi, Salerno
Italy

Pietro G. Morasso
University of Genova
Dipartimento di Informatica,
Sistemistica, Telematica
Via Opera Pia 11A
16145 Genova
Italy

ISBN 3-540-19887-3 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-19887-3 Springer-Verlag New York Berlin Heidelberg

Printed in Great Britain

Typesetting: Camera ready by contributors
Printed by Athenæum Press Ltd., Newcastle upon Tyne
34/3830-543210 Printed on acid-free paper

Contents

Bayesian Properties and Performances of Adaptive Fuzzy Systems in Pattern Recognition Problems

F. Masulli[1,3], F. Casalino[2], and F.Vannucci[3]
(1) Department of Physics - University of Genoa
Via Dodecaneso 33 - 16146 Genova (Italy)
(2) DISI - Department of Computer and Information Sciences
University of Genoa - Via Benedetto XV, 3 - 16132 Genova (Italy)
(3) INFM Research Unit of Genoa
Via Dodecaneso 33 - 16146 Genova (Italy)

1 Introduction

Generally, Neural Networks are used to solve problems for which a-priori knowledge is provided, in an implicit way, through numerical relationships among variables (e.g., pattern recognition). Fuzzy Systems are successfully employed mainly to solve problems for which a-priori knowledge is available in linguistic form (e.g., process control).

As will be shown in this article, Adaptive Fuzzy Systems (AFS) [1] can be used to handle problems for which a-priori knowledge is available only in numerical form, or for which it could be too expensive to render knowledge explicit in linguistic form.

Multi-Layer Perceptrons (MLP) and Adaptive Fuzzy Systems can be trained, in a supervised way, to map variables without using explicit hypotheses about the analytical dependences among them. Such methods are usually referred to as function approximators (*model-free estimators*) [2, 3].

Therefore, both types of models can be applied in various domains (e.g., signal processing, process control, pattern recognition, etc.), and the criteria for choosing between them for a given application are not yet clear.

In this paper, we shall use an Adaptive Fuzzy System to solve a pattern recognition problem, i.e. off-line recognition of handwritten characters. We shall demonstrate that the AFS approximates a Bayesian discriminant function; moreover, we shall experimentally verify that, in the training phase, this system is some order of magnitude faster than a Multi-Layer Perceptron.

2 The Adaptive Fuzzy System

Fuzzy sets, proposed by Zadeh in 1965 [4], can be defined through a membership function m_F that maps the elements of the universal set in the unit range $[0, 1]$.

The form of the membership function is arbitrary. In this way, it is possible to model set of objects that fulfil a given property in different ways.

Fuzzy rules are expression of the type:

if A then B

where A and B are labels associated with fuzzy sets that are characterized by suitable membership functions.

An Adaptive Fuzzy System (AFS) [1] is a feedforward system that could be regarded as a Multi-Layer Perceptron with only one hidden layer; the units of the MLP correspond to fuzzy rules.

Let us describe our AFS implementation.

If there are K units in the input layer, J rules in the hidden layer and I units in the output layer, the activation of the j-th rule can be expressed as:

$$R_j = \prod_k \mu_{jk}(x_k),$$

where the quantity $\mu_{jk}(x_k)$ is the value of the membership function of the component x_k of the input vector for the j-th rule.

The membership function can be defined as:

$$\mu_{jk}(x_k) = \exp(-\frac{|x_k - m_{jk}|^2}{2\sigma_{jk}^2}).$$

Therefore, the receptive fields overlap with one another.

The values of the output units are obtained by means of defuzzification process based on the centroid rule [2] :

$$y_i = \frac{\sum_j R_j s_{ij}}{\sum_j R_j}.$$

The AFS is trained to work as a classifier by minimizing the Mean Square Error between the output of the net and the label vector $\vec{\mu}$, whose components are defined as follows:

$$\mu_j = \begin{cases} 1 & \text{if the example belongs to the class } j \\ 0 & \text{otherwise.} \end{cases}$$

The learning formulas for the the system parameters (i.e. m_{jk}, σ_{jk} and s_{ij}) are obtained by the Back-Propagation technique [1].

In our implementation, we make the learning rates adaptive, thus obtaining a sharp reduction in the convergence time, as compared with fixed learning rates [5].

It is worth noting that a classification approach using an Adaptive Fuzzy System allows us to insert available a-priori knowledge in the rules before the training phase and to interpret the learned values of parameters in terms of rules.

3 The AFS as an Approximation to a Bayes Optimal Discriminant Function

As stated in the Introduction, an AFS can approximate functions [2]. So it is possible to demonstrate that the AFS can approximate the Bayesian optimal discriminant function [6], if a large training set is used. This can be easily accomplished on the basis of the demonstration developed by Ruck et al for a Multi-Layer Perceptron [7].

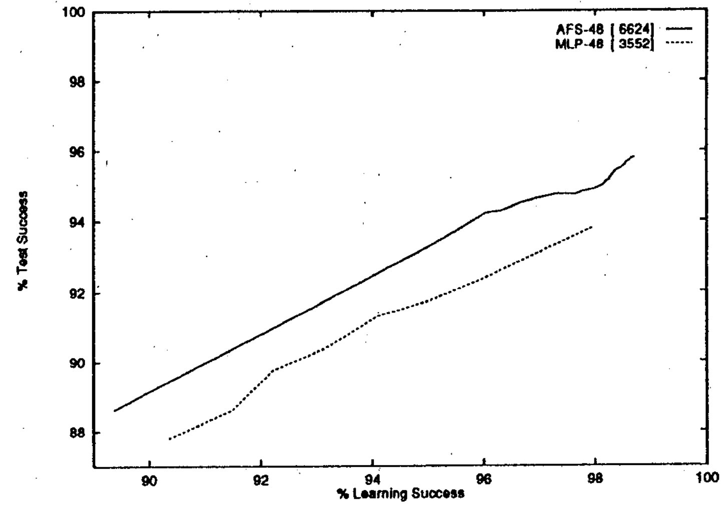

Figure 1

rules	parameters	$(\%L, \%T)$	epochs	(%L1, %T1)	Time (min)
24	3312	(97.55, 94.00)	30	(91.0, 88.0)	3.5
32	4416	(98.05, 95.00)	30	(89.0, 87.0)	4.0
48	6624	(98.65, 95.81)	30	(91.5, 88.5)	16.0
64	8832	(98.32, 96.02)	20	(93.5, 92.0)	23.0
128	17664	(98.97, 96.20)	30	(94.5, 93.5)	46.0

Table 1

4 Data Set and Preprocessing

The samples of the training and the test sets were extracted from the NIST-3 CD-ROM [8], which includes 313389 segmented characters in 128×128 binary image format and their corresponding labels in ASCII format.

During the preprocessing phase, the current character is normalized to a 32×32 format. A low-pass filter is used to fill small holes and remove some small noisy spots. Then a shear transformation is applied to the character. Finally, the dimensions of the input space are further reduced so that each character may be represented as a 64-component vector. Each component is associated with the number of black pixels in 4×4 disjoint image subsquares.

The resulting data-base contains a learning set and a test set, both made up of $10,000$ decimal numerals.

5 Results and Discussion

In Figure 1, a comparison between the performances of the Multilayer Perceptron (MLP) and of the Adaptive Fuzzy Systems (AFS) is reported. The numbers within brackets refer to the parameters used by the two networks.

The MLP and the AFS exhibit similar generalization accuracies. This fact can be theoretically explained by the two networks' capabilities for approximating the Bayesian optimal discriminant function, as previously discussed in Section 3.

Table 1 gives the number of rules, the number of adaptive parameters, the percentages of learning success (%L) and of test success (%T), the number of epochs required by the training phase, the percentages of learning success (%L1) and of test success (%T1), and the duration (in min) of each epoch. Some results are quite impressive: e.g., in the case of an AFS with 128 rules, a single epoch of 46 min, is enough to reach a test success %T1 = 93.5.

6 Acknowledgments

This work was supported by grants from CNR-Progetto Strategico Reti Neurali, GNCB-CNR, Consorzio INFM and MURST. Part of this work was carried out in Summer 1993 while F. Masulli was a Senior Visiting Scientist at the International Computer Science Institute in Berkeley (USA). We thank Alessandro Sperduti and Maurizio Martelli for helpful discussions.

References

[1] C.C. Jou, "On the mapping capabilities of fuzzy inference systems", *in IJCNN International Joint Conference on Neural Networks*, pp. 703–713, Baltimore, MD, USA, 7-11 June 1992, 1992. IEEE, New York, NY.

[2] B. Kosko, editor, *Neural networks and fuzzy systems : a dynamical systems approach to machine intelligence*, Englewood Ciffs Prentice Hall, NJ, 1992.

[3] K. Funahashi, "On the approximation realization of continuous mappings by neural networks", *Neural Networks*, vol. 2, pp. 183–192, 1989.

[4] L.A. Zadeh, "Fuzzy sets", *Information and Control*, vol. 8, pp. 338–352, 1965.

[5] F. Casalino, "Fuzzy systems for handwriting recognition (in italian)", Laurea thesis in computer science, University of Genoa, Genoa - Italy, 1993.

[6] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

[7] D.W. Ruck, S.K. Rogers, M. Kabrisky, M.E. Oxley, and B.W. Suther, "The multilayer perceptron as an approximation to a bayes optimal discriminant function", *IEEE Transactions on Neural Networks*, vol. 1, pp. 296–298, 1990.

[8] M.D. Garris and R.A. Wilkinson, *NIST Special Database3 Handwritten Segmented Characters*, National Institute of Standard and Technology, Gaithesburg, MD , USA, 1992.