PROCEEDINGS REPRINT

SPIE—The International Society for Optical Engineering

Reprinted from

Applications of Fuzzy Logic Technology III

10–12 April 1996
Orlando, Florida

SPIE PROCEEDINGS SERIES

Volume 2761

Fuzzy Systems in High Energy Physics

Marcello Castellano

Istituto Nazionale di Fisica Nucleare
Sezione di Bari
Via Amendola 173 - 70126 Bari - Italy
email : castellano@bari.infn.it

Francesco Masulli and Massimo Penna

Istituto Nazionale di Fisica della Materia
Dipartimento di Fisica - Università di Genova
Via Dodecaneso 33 - 16146 Genova - Italy
email : {masulli| penna}@ge.infm.it

Abstract

Decision making is one of the major subject of interest in physics. This is due by the intrinsic finite accuracy of measurement that leads to the possible results to span a region for each quantity. In this way, to recognize a particle type among the others by a measure of a feature vector, a decision must be made. The decision making process becomes a crucial point whenever a low statistical significance occurs as in space cosmic ray experiments where searching in rare events requires to reject as many background events as possible (high purity), keeping as many signal events as possible (high efficiency). In the last few years, interesting theoretical results on some feedforward connectionist systems (FFCSs) have been obtained. In particular, it has been shown that multilayer perceptrons (MLPs), radial basis function networks (RBFs), and some fuzzy logic systems (FLSs) are nonlinear universal function approximators. This property permits to build system showing intelligent behaviour, such as function estimation, time series forecasting, and pattern classification, and able to learn their skill from a set of numerical data. From the classification point of view, it has been demonstrated that non-parametric classifiers based FFCSs holding the universal function approximation property, can approximate the Bayes optimal discriminant function and then minimize the classification error. In this paper has been studied the FBF when applied to a High Energy Physics problem. The FBF is a powerful neuro-fuzzy system (or adaptive fuzzy logic system) holding the universal function approximation property and the capability of learning from examples. The FBF is based on product-inference rule (P), the Gaussian membership function (G), a singleton fuzzifier (S), and a center average defuzzifier (CA). The FBF can be regarded as a feedforward connectionist system with just one hidden layer whose units correspond to the fuzzy MIMO rules. The FBF can be identified both by exploiting the linguistic knowledge available (structure identification problem) and by using the information contained in a data set (parameter estimation problem). The fuzzy system has been found to be effectiveness for the classification tasks of about $2 \mathrm{x} 10^{-3}$ hadron contamination at 90% of electron acceptance. A comparison between the adaptive system results and the others previous one obtained by using both statistical and neural network based methodologies will be also presented.

Keywords: High Energy Physics, Classification, Fuzzy Systems, Fuzzy Basis Functions.

0-8194-2142-1/96/$6.00

1 INTRODUCTION

Decision making is one of the major subjects of interest in physics. For instance, to recognize a particle type among others by using the measure of a feature vector, a decision must be made. The decision making process becomes a crucial point in the case of a low statistical significance, as in space cosmic ray experiments where searching for rare events requires that as many background events as possible be rejected (high purity), keeping as many signal events as possible (high efficiency).

In the last few years, interesting theoretical results on some feedforward connectionist systems (FFCSs) have been obtained. In particular, it has been shown that multilayer perceptrons (MLPs), radial basis function networks (RBFs), and some fuzzy logic systems (FLSs) are universal function approximators.[1–2] This property make it possible to build systems showing intelligent behaviours, such as function estimation, time series forecasting, and pattern classification.[3] From the classification point of view, it has been demonstrated that non-parametric classifiers based on FFCSs with the universal function approximation property, can approximate the Bayes optimal discriminant function.[4–5]

In this paper, we report preliminary results on the study of the classification power of a classifier based on a FBF, when applied to discriminate signal/background events with high purity and high efficiency in High Energy Physics. The data set used was collected at the CERN-PS exposing a transition radiation detector to an electron/hadron test beamm. The FBF parameters extimation was performed by a gradient descent with respect to the MSE across the training set. In particular, for the data set described in Section 3, we select the optimal structure and rejection threshold for a classifier based on neuro-fuzzy system constituted by a network of *Fuzzy Basis Functions* (FBF's), previously presented in[6–8] in order to minimize the *contamination*, when a reasonable of *acceptance* of the classifier is assigned [1].

In the next section the fuzzy logic system is presented. In Section 3, we describe the experimental data set. In Section 4, the learning procedure and the problem of the rejection are presented. In Section 5, the results are reported and discussed. Conclusions are presented in Section 6.

2 THE FBF NETWORK

In the application reported in this paper, we used a FBF network[6,7] based on the following assumptions: height method defuzzifier, product-inference rule, singleton fuzzifier, and Gaussian membership function. Specifically, if there are K units in the input layer, J fuzzy inference rules and I outputs, the rule activations can be written as:

$$r_j = \prod_i \mu_{ji}(x_i), \tag{1}$$

The quantity $\mu_{ji}(x_i)$ represents the value of the membership function of the component x_i of the input vector for

[1]The acceptance and the contamination are defined as:

$$\text{acceptance} = \frac{\text{output signal}}{\text{input signal}}$$

$$\text{contamination} = \frac{\text{output background}}{\text{output signal}}$$

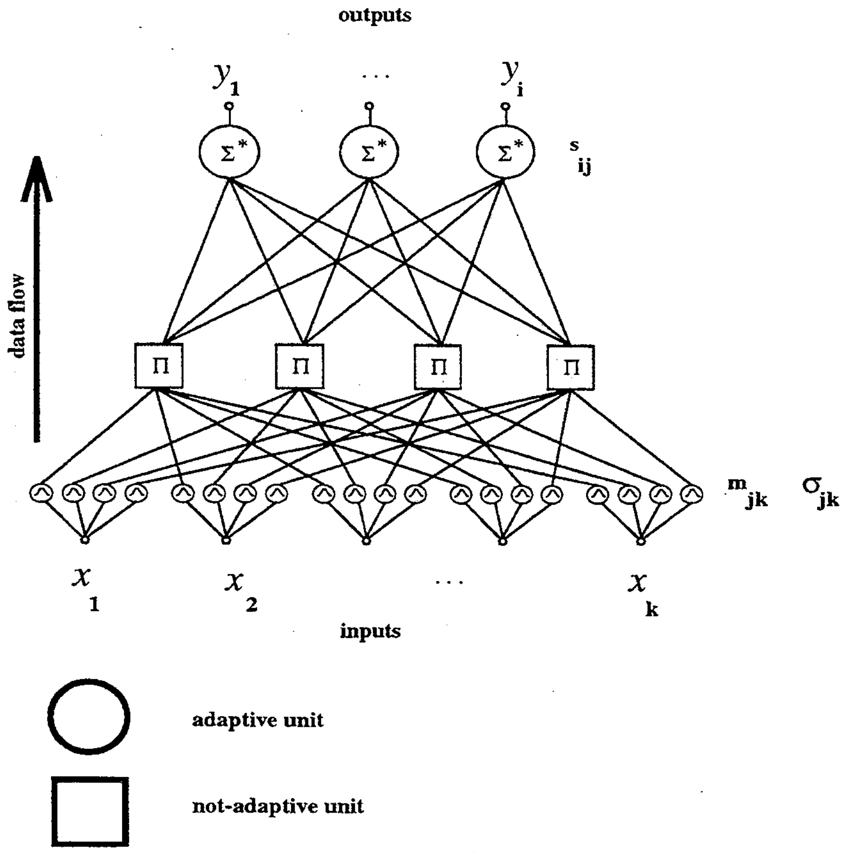


Figure 1: Connectionist interpretation of the FBF network.

the j-th rule, and is defined as:

$$\mu_{ji}(x_i) = \exp\left(-\frac{(x_i - m_{ji})^2}{2\sigma_{ji}^2}\right), \tag{2}$$

where m_{ji} and σ_{ji}^2 are the means and the variances. The values of the output units are:

$$y_k = \frac{\sum_j r_j s_{kj}}{\sum_j r_j}, \tag{3}$$

and s_{ij} is the maximum value of the output fuzzy membership function of the $j-th$ rule associated with the output y_i. Without loss of generality, we assume that the fuzzy membership functions are singletons.

The FBF network can be regarded as a feedforward connectionist system with one hidden layer whose units correspond to the fuzzy rules. In Figure 1 a connectionist interpretation of the FBF network is shown. In the drawing, circles represent adaptive units, while squares stand for not-adaptive units, and free parameters are evidenced near the corresponding units.

In,[9,7] on the basis of the Stone-Weierstrass Theorem[10] the Universal Approximation Theorem was demonstrated that guarantees that a FBF network can perform function approximation at an assigned precision. As is well known, similar results on function approximation have been obtained by other feedforward connectionist systems, such as MLP's and RBF networks.[1,11]

The FBF network can be identified both by exploiting the linguistic knowledge available (*structure identification problem*)[12] and by using the information contained in a data set (*parameter estimation problem*).[12]

A non-parametric classifier able to approximate the Bayes discriminant function, can be realized by training a FBF network by an algorithm able to find the values of the parameters (or *weights*) that minimize a suitable cost function, like the *mean square error* (MSE):

$$MSE = \frac{\sum_{i,n}(y_i^n - t_i^n)^2}{N} \quad , \tag{4}$$

where N is the size of the training set, $\mathbf{y}^n = (y_i^n)$ is the network output, and $\mathbf{t}^n = (t_i^n)$ is the n-th label of the associative pair of the training set. The components of $\mathbf{t}^n$ are defined as follows:

$$t_i = \begin{cases} 1 & \text{if the pattern belongs to class } i, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The cost function (4) can be minimized by many different techniques, among which the gradient descent technique, clustering methods,[7] Kalman filters,[13] genetic algorithms,[14] , etc. In our application, the FBF network parameters (i.e., m_{jk}, σ_{jk} and s_{ij}) were obtained by performing a gradient descent with respect to the MSE across the training set. The learning formulas are as follows:[6,7]

$$\Delta s_{ij} = \eta_s [t_i - y_i]\psi_j \tag{6}$$

$$\Delta m_{jk} = \eta_m \psi_j \sum_i [t_i - y_i][s_{ij} - y_i][x_k - m_{jk}]/\sigma_{jk}^2 \tag{7}$$

$$\Delta \sigma_{jk} = \eta_\sigma \psi_j \sum_i [t_i - y_i][s_{ij} - y_i][x_k - m_{jk}]^2/\sigma_{jk}^3 \tag{8}$$

where the *fuzzy basis functions*[15]

$$\psi_j = \frac{\prod_k \mu_{jk}(x_k)}{\sum_j \prod_k \mu_{jk}(x_k)} \tag{9}$$

represents the normalized activation of the rule j, and η_s, η_m, and η_σ are the learning rates of s_{ij}, m_{jk}, and σ_{jk}, respectively.

If a linguistic description of the classes considered is available, in addition to the numerical training set, the FBF network permits one to integrate the two kinds of information. On the contrary, if a linguistic description is not available, as in our application, the structure identification must be achieved experimentally according to a performance-based criterion.

	Data set	Training set	Test set	Validation set
Events	3000	1500	500	1000
Signal (Electrons)	1849	930	307	612
Background (Pions)	1151	570	193	388
Signal/Background	1.61	1.63	1.59	1.58

Table 1: Data distribution.

3 EXPERIMENTAL DATA SET

We tested the classification properties of the FBF network in an application addressed to events discrimination with high purity and high efficiency in High Energy Physics.

The data set was collected at the CERN-PS by exposing a transition radiation detector to an electron/hadron test beam, with a momentum of 4.0 GeV/c. We used a labeled data set of 3000 events, represented as vectors of 10 components in the interval $[0, 1]$. The data set was subdivided into (see Table 1):

- a training set of 1500 events, to be used for the estimation of the system parameters;
- a test set of 500 events, to be used to implement the stop criterion of the training process;
- a validation set of 1000 events, for evaluating the system performances after the learning.

4 METHODS

We used the FBF network in order to obtain an optimal classifier able to minimizie the *contamination* particular, for the data set described in Section 3, we at a level of *acceptance* set to the value of 90%. We used a FBF network with 10 inputs and two outputs, corresponding to the signal (electrons) and the background (pions) classes. In the next section we will show how the number of rules was chosen on the basis of performance criteria.

For each configuration of the fuzzy system, the free-parameters where initializated at random, and then the training was performed *by pattern*,[3] i.e., by changing the parameter values, following the gradient descent algorithm, after the presentation of each example.

After each *epoch* (i.e., the presentation of the full training set), the MSE across the test set was evaluated. The training procedure was stopped when the minimum of MSE was reached, otherwise an new epoch was performed.

Note that, in order to speed up the training phase, an adaptive learning-rate scheme[16] was adopted, as proposed by Vogl et al.[17]

After training the fuzzy system can recognize an event as belonging to one of the two classes (Signal and Background).

The generalization capabilities on the validation set are then studied in order to set a *rejection threshold* regulating the relation between contamination and acceptance.

Rules	2	3	5	9	13	18	22
Acceptance ($\cdot 10^{-2}$)	92.97	90.69	92.16	91.34	93.46	91.18	92.81
Contamination ($\cdot 10^{-3}$)	3.51	1.80	3.55	1.79	3.50	1.79	3.52

Table 2: Best results of with different FBF network structures and a rejection threshold of .7.

If the difference between the values of the two outputs exceed the assigned value of the rejection threshold, the event is recognized as belonging to the class corresponding to the output with the larger value; otherwise, the event is rejected (i.e. it is not sent to any classifier output).

5 RESULTS AND DISCUSSION

In order to evaluate the optimal number of rules to be used for the FBF network, many trials (each of them consisting in 10 independent tests) were performed using structures of different number of rules and rejection thresholds. In Table 2, the best result for each structure of the fuzzy system at the rejection threshold of .7 are presented.

It is worth noting that, due to the low statistics, marked differences among the considered structures cannot be found.

We chose the structure with 3 rules, after analyzing the best results reported in Table 2, the averaged results over the ten tests, and considering that a system with few parameters guarantees the best generalization.

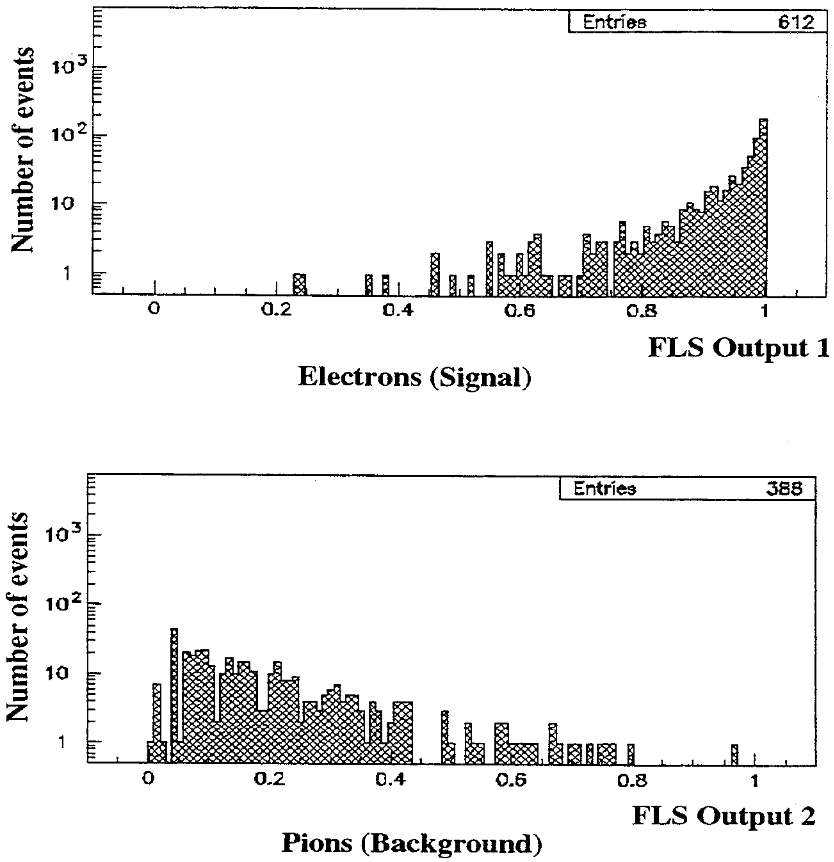
In Table 3, for a FBF network using three rules and an acceptance value above 90%, the best contamination results (on trials of 10 independent tests) for different values of the rejection threshold, are reported.

In Figure 2, for the optimal FBF network (i.e., a FBF network using three rules and a rejection threshold value of .7), the distributions of the output values (*feature space*) for the two classes are given.

In Figure 3, for the optimal FBF network the contamination is plotted versus the acceptance.

Threshold	0.3	0.4	0.5	0.6	0.7	0.8
Acceptance ($\cdot 10^{-2}$)	90.03	90.36	92.48	90.69	90.69	90.52
Contamination ($\cdot 10^{-3}$)	1.81	1.81	3.53	1.80	1.80	1.81

Table 3: Performances of the FBF network using three rules and different values for the rejections threshold.



:put distributions on the validation set, obtained by a FBF network using 3 rules and the rejection
ıe .7. Feature space for the 10-input fuzzy logic architecture.

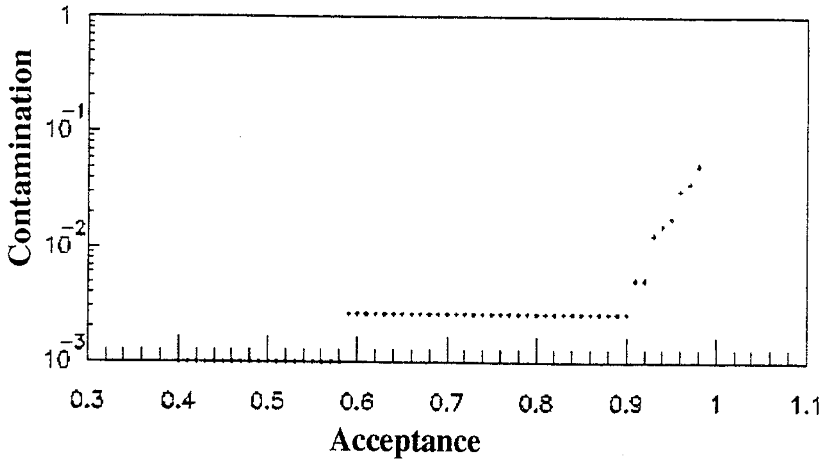


Figure 3: Contamination versus acceptance for the fuzzy system of Figure 2.

6 CONCLUSIONS

A FBF network has been applied for the signal/background classification problem in High Energy Physics. Its ability has been assessed on a data set collected at the CERN-PS by exposing an imaging detector to an electron/hadron test beam. The fuzzy system has been found to be effective for the classification tasks, as it has shown a low degree of background contamination at high signal acceptance. This result suggests that the fuzzy approaches can be employed in High Energy Physics experiments to search for rare events with high accuracy.

7 ACKNOWLEDGEMENTS

This work was supported by grants from INFM, INFN, and MURST. We thank Franco Casalino for helpful discussions.

8 REFERENCES

[1] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2:303–314, 1989.

[2] J.M. Mendel. Fuzzy logic systems for engineering: A tutorial. *Proceedings of the IEEE*, 83:345–377, 1995.

[3] S. Haykin. *Neural Networks : a Comprehensive Foundation.* Macmillan, New York, Toronto, 1994.

[4] 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*, 1:296–298, 1990.

[5] F. Casalino, F. Masulli, A. Sperduti, and F. Vannucci. Semantic phase transition in a classifier based on an adaptive fuzzy system. In *Proceedings of the Third IEEE International Conference on Fuzzy Systems, IEEE-FUZZ94*, volume 2, pages 808–812, Orlando, FL, USA, 1994. IEEE.

[6] C.C. Jou. Comparing learning performance of neural networks and fuzzy systems. In *IEEE International Conference on Fuzzy Systems*, pages 1028–1033, San Francisco, 1993. IEEE, New York, NY.

[7] L. X. Wang. *Adaptive Fuzzy Systems and Control.* Prentice Hall, Englewood Cliffs, New Jersey, 1994.

[8] F. Masulli, F. Casalino, and F. Vannucci. Bayesian properties and performances of adaptive fuzzy systems in pattern recognition problems. In M. Marinaro and P.G. Morasso, editors, *Proceedings of the European Conference on Artificial Neural Networks, ICANN-94*, pages 189–192, Sorrento, Italy, 1994. Springer.

[9] C.C. Jou. On the mapping capabilities of fuzzy inference systems. In *IJCNN International Joint Conference on Neural Networks*, volume 2, pages 703–713, Baltimore, MD, USA, 1992. IEEE.

[10] W. Rudin. *Principles of Mathematical Analysis.* McGraw-Hill, Inc., New York, 1960.

[11] T. Poggio and F. Girosi. Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247:978–982, 1990.

[12] C.C. Lee. Fuzzy logic in control systems: fuzzy logic controller. I. *IEEE Transactions on Systems, Man and Cybernetics*, 20:404–418, 1990.

[13] J.S.R. Jang. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. on Systems, Man, and Cybernetics*, 23:655–684, 1993.

[14] R. Caviglia. Soft-computing methods for time series forecasting. Laurea Thesis in Computer Science, University of Genoa, Genova - Italy, 1994. In Italian.

[15] L. Wang and J.M. Mendel. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Trans. on Neural Networks*, 5:807–14, 1992.

[16] F. Casalino. Fuzzy systems for handwriting recognition. Laurea Thesis in Computer Science, University of Genoa, Genova - Italy, 1993. In Italian.

[17] T.P. Vogl, J.K. Mangis, A.K. Rigler, W.T. Zink, and D.L. Alkon. Accelerating the convergence of the back-propagation method. *Biological Cybernetics*, 59:257–263, 1988.