RATIONAL APPROXIMANT ARCHITECTURE FOR NEURAL NETWORKS

F.M. Frattale Mascioli & G. Martinelli
Dip. INFO-COM, Università di Roma "La Sapienza"
via Eudossiana, 18 - 00155 Roma - Italy
Tel: +39 6 44585488 / 9; fax: +39 6 4873300
e-mail: mascioli@infocom.ing.uniroma1.it

ABSTRACT

A novel approach is proposed for overcoming the multiple minima problem, present in the learning of a supervised neural network. It allows to connect rational function approximations to neural networks and is based on the use of a truncated Fourier expansion for determining: 1) the architecture; 2) the parameters of the net, avoiding local minima in an efficient way.

1 INTRODUCTION

When determining a neural network for realizing an unknown mapping through examples, a very difficult problem is to overcome the multiple minima present in the objective function. A possible solution is to use a unimodal objective function. However, this solution usually yields poor approximations of the mapping, since it does not rely on an efficient algorithm. In the present work we follow for this purpose a recent approach which allows to connect rational function approximations to neural networks [1]. The connection regards the class of neural networks constituted by a single hidden layer of N sigmoidal neurons, as shown in Fig. 1.

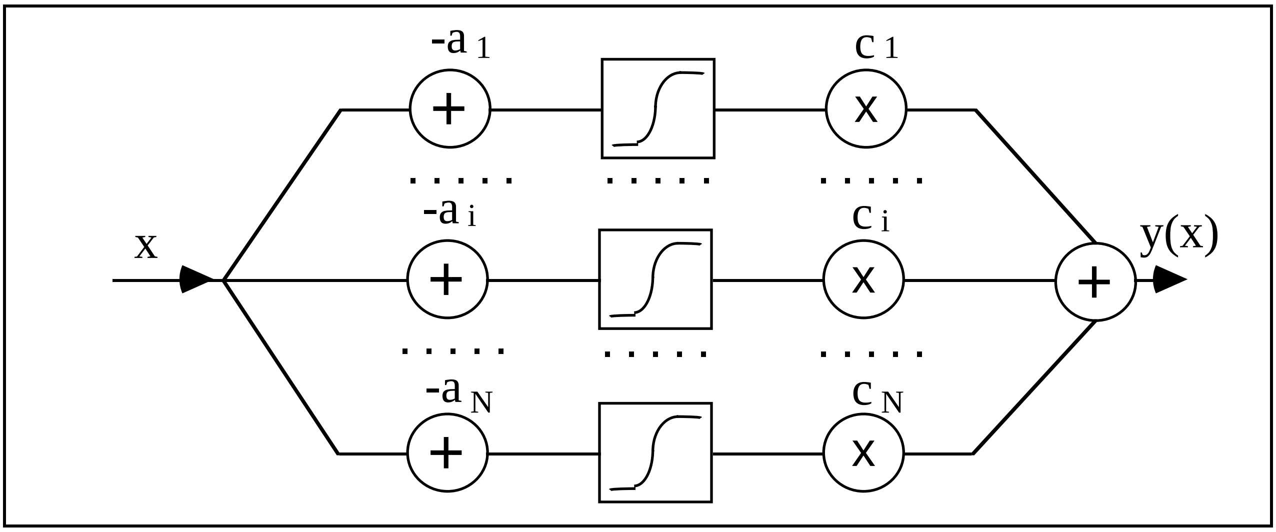

Fig. 1: Class of neural networks considered in [1]. The box represents a sigmoidal activation function.

The output of this network is given by

(1) $$y(x)=\sum_{i=1}^{N} c_i \sigma(x-a_i) \ ,$$

where:

$$\sigma(x)=(1+e^{-x})^{-1}, \quad a_i \neq a_j \quad for\ i \neq j \ .$$

In terms of a transformed input variable

(2) $$z=e^{x},$$

the output of the neural network becomes a rational function, i.e.:

(3) $$F(z) = z \cdot H(z),$$

with:

$$H(z) = \sum_{i=1}^{N} \frac{c_i}{z + \alpha_i} \text{ and } \alpha_i = e^{a_i} .$$

Function $H(z)$ is characterized by positive real distinct poles; the synapses c_i coincide with their residues.

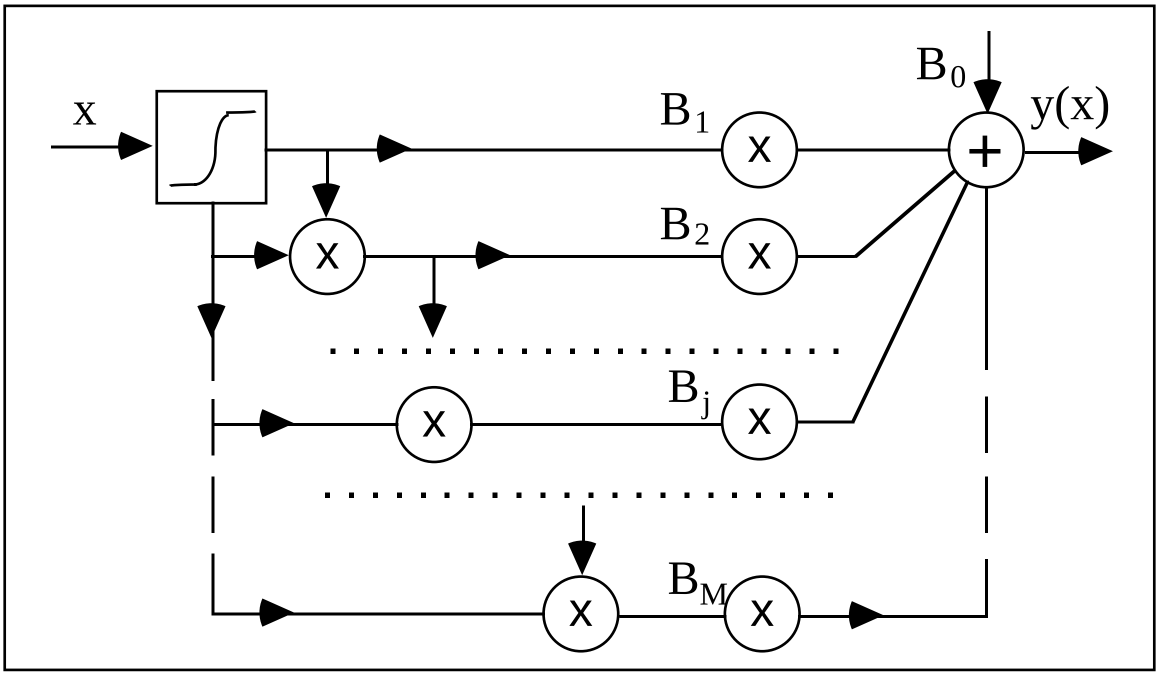

Fig. 2: Neural network based on Fourier rational approximant.

2 THE PROPOSED ARCHITECTURE

The neural network of Fig. 1 only realizes a limited class of rational functions. In fact, optimal rational approximants of a given function may possess multiple complex poles. Consequently, more general schemes of neural networks should be provided in order to take advantage of the said connection. To any specific algorithm available for the rational approximation, a particular architecture corresponds.

We will illustrate this correspondence by applying an algorithm based on Fourier series expansion. The result will be a novel architecture for the neural network.

Let f(x) $\{x, f(x) \in R\}$ be the function to be approximated. By replacing x for

(4) $$w = 2\tan^{-1}\sqrt{e^{x}}$$

function f(x) becomes a function g(w) defined in the range 0÷π. Outside this range, it is convenient to consider g(w) as an even periodic function with period equal to 2π. Its Fourier series, truncated to the M-th harmonic, is equal to

(5) $$g(w) \cong \sum_{k=0}^{M} A_k \cos kw .$$

We note that:

(6) $$\begin{aligned} \cos kw &= \cos 2k \cdot \tan^{-1}\sqrt{z} = \\ \cos k \cdot \sin^{-1}\left(\frac{2\sqrt{z}}{1+z}\right) &= \\ \cos k \cdot \cos^{-1}\left(\frac{1-z}{1+z}\right) &= \\ C_k\left(\frac{1-z}{1+z}\right) &= \sum_{i=0}^{k} C_{ki}\left(\frac{1-z}{1+z}\right)^{i} \end{aligned}$$

having introduced the Chebyshev polynomial $C_k(.)$ of order k. Hence, by using (6) and taking account of (2), formula (5) becomes:

$$\text{(7)} \qquad G(z) = \sum_{j=0}^{M} B_j \left(\frac{z}{1+z} \right)^j ,$$

with

$$B_j = (-2)^j \sum_{k=j}^{M} A_k \sum_{i=j}^{k} C_{ki} \binom{i}{j} \; .$$

Function G(z) is the resulting rational approximant. The values B_j are obtained from a simple Fourier series expansion and from the coefficients of Chebyshev polynomials, available in any mathematical handbook. Formula (7) corresponds to the neural network of Fig. 2, which simply coincides with a linear combiner driven by the successive powers of the input, predistorted by passing through a neuron having a sigmoidal activation function.

The learning of this network is without difficulty, since it can be based on a quadratic error function. Moreover, the resulting network is efficient since it is based on Fourier expansion.

3 REFERENCE

[1] R.C. Williamson & U. Helnike: *"Existence and uniqueness results for neural network approximation"*, IEEE Trans. on Neural Networks, 6, 1995, pp. 2-13.