Π–DTB, Discrete Time Backpropagation with Product Units

J. Santos and R. J. Duro
Grupo de Sistemas Autónomos - Departamento de Computación - Univ. de A Coruña
Facultad de Informática - Campus de Elviña - A Coruña 15071 Spain
{ santos, richard }@udc.es

Abstract

This paper introduces a neural network that combines the power of two different approaches to obtaining more efficient neural structures for processing complex real signals: the use of trainable temporal delays in the synapses and the inclusion of product terms within the combination function. In addition to the neural network structure itself, the paper presents a new algorithm for training this particular type of networks and provides a set of examples using chaotic series, which compare the results obtained by these networks and training algorithm to other structures.

1. Introduction

Even though the term “high order” is not clearly delimited in the artificial neural network field, it encompasses those networks that incorporate elements with a larger processing capacity than the classical nodes, that use linear or non-linear activation functions acting over a combination that is the addition of the contribution of each input modulated by the corresponding connection weight.

Giles and Maxwell [4] define the order of a neural network through an interpretation of the equation:

\[ y_i(x) = S \left[ w_0(i) + \sum_j w_1(i,j)x(j) + \sum_j \sum_k w_2(i,j,k)x(j)x(k) + \ldots \right] \]

That is, in addition to the summation term of first order units, the higher order units use both addition and multiplication. Through these processes they define the so called Sigma-Pi units. Consequently, and using this definition, a unit that includes terms up to and including degree \( k \) will be called a \( k \text{-th-order} \) unit. For these networks to be useful, as indicated by Giles and Maxwell [4], they should be matched to the order of the problem, and in this sense, a single \( k \text{-th-order} \) unit will solve a \( k \text{-th-order} \) problem, as defined by Minski and Papert [8].

Many approaches have been used to obtain non-linearity in the nodes [5], some even using the conditional operator leading to the concept of inferential neural networks studied by Mira [9], where one implementation can be found in [10]. The authors have employed a symbolic formalism (a micro-frame based model) to represent the information processing and learning algorithms of the connectionist models.

Notwithstanding the previous comments the increase in the processing power of a neural network cannot only lay in the capacity of the nodes. The synaptic connections of a neural network are forgotten elements that can also be used to this end. Two
examples of this can be found in [2] and [3]. The first one considers neural networks that include a variable delay term in each synapse in addition the connection weight. It represents the time the information requires to go from the origin node to the postsynaptic node and resembles the length of the biological synapses. Like the weights, the time delays between the nodes are trainable. The second case [3] corresponds to substituting the numerical weight term in the synapses by a gaussian function with three free trainable parameters: amplitude, variance and center of the gaussian. This way, the contribution of a connection to the postsynaptic node is not only a function of the connection weight, but also of the current value of the input to that connection. In other words, the gaussian functions act as a filter of the inputs, which is automatically obtained though learning to the best adaptation to the problem. These structures demonstrate their higher processing capabilities over traditional ANN architectures with a reduction in the number of processing elements.

Particularizing in the application of this paper, in the realm of signal identification and modelling, the main problem is to obtain a model of the signal that is as accurate as possible and which permits performing short and long term predictions. This is, we are interested in the dynamic reconstruction of the measured signal.

Most neural network based implementations have not taken into account the basic premises on which the modelling of signals must be based, that is, that in order to model a time dependent non linear signal one must be able to define its state space, or an equivalent state space, in an unambiguous manner in order to prevent multivalued sets of coordinates which would lead to ambiguous predictions. In fact, one of the best areas to appreciate this is in the case of chaotic time series, where if the state space is not well chosen the orbits of the series in this space will cross, generating points of ambiguity. In this article we make use of a trainable delay based artificial neural network that directly implements a form of the embedding theorem [6][11], with the advantage of being able to autonomously obtain the embedding dimension and the normalized embedding delay.

As a first step in directly training the delays introduced in a generalized synaptic delay network, using a gradient descent method, we have developed the DTB algorithm [2], that results in easy and precise training of temporal event processing networks straight from the input signals, with no windowing or preprocessing required, and what is more important, the embedding dimension does not have to be defined a priori. In the context of signal processing, and taking into account that many signal related correlations take place in the frequency domain, we have now added the possibility of including product terms in some of the nodes which, combined with the automatic selection of the signal points to be multiplied (through the training of the delay terms), allow for learnable signal correlation operations within the network.

In the following sections we will present the network and algorithm used to train it as well as some examples of its application to chaotic signal modelling.

2. Discrete Time Backpropagation with Π Units

The artificial neural network we consider for training consists of several layers of neurons connected as a Multiple Layer Perceptron (MLP). There are two differences with traditional MLPs. The first one is that the synapses include a delay term in addition to the classical weight term. That is, now the synaptic connections between