NOVEL COMPLEX VALUED NEURAL NETWORKS

Garimella Rama Murthy,
Associate Professor,
International Institute of Information
Technology (IIIT), Hyderabad, India
E-mail: rammurthy@iiit.ac.in

Narendra Ahuja,
Distinguished Professor
University of Illinois at Urbana-Champaign
(UIUC), Urbana, Illinois, USA

ABSTRACT

In view of many applications, in recent years, there has been increasing interest in complex valued neural networks. In this paper, it is reasoned that transforming real valued signals into complex valued signals (using Discrete Fourier Transform) and processing them in that domain is equivalent to processing real valued signals. This approach could have many advantages. Also neural networks based on a novel model of neuron are proposed. Some interesting open questions are proposed.

I. INTRODUCTION

Starting in 1950's researchers tried to arrive at models of neuronal circuitry. Thus the research field of artificial neural networks took birth. The so called, perceptron was shown to be able to classify linear separable patterns. Since the Ex-clusive OR gate cannot be synthesized through any perceptron (as the gate outputs are not linearly separable), the interest in artificial neural networks faded away. In the 1970's, it was shown that multi-layer feed forward neural network such as a multi-layer perceptron is able to classify non-linearly separable patterns.

Living systems/machines such as homosapiens, lions, tigers etc have the ability to associate externally presented one/two/three dimensional information such as audio signal/images/three dimensional scenes with the information stored in the brain. This highly accurate ability of association of information is amazingly achieved through the bio-chemical circuitry in the brain. In 1980's Hopfield revived the interest in the area of artificial neural networks through a model of associative memory. The main contribution is a convergence theorem which shows that the artificial neural network reaches a memory/stable state starting in any arbitrary initial input (in a certain important mode of operation). He also demonstrated several interesting variations of associative memory. In (Rama4), a continuous-time version of associative memory is described. It is shown that the celebrated convergence Theorem in discrete time generalizes to the continuous time associative memory. In (Rama2), the model of associative memory in one dimension (Hopfield associative memory) is generalized to multi/infinite dimensions and the associated convergence theorem is proven.

It was realized by researchers such as N.N Aizenberg that the basic model of a neuron must be modified to account for complex valued inputs, complex valued synaptic weights and thresholds [AAV]. In many real world applications, complex valued input signals need to be processed by neural networks with complex synaptic weights [Hir]. Thus the need to study, design and analysis of such networks is real. Also, in (Rama3) the results on real valued associative memories are extended to complex valued neural networks. In [Nit1, Nit2], the celebrated back propagation algorithm is generalized to complex valued neural networks. Also, in [Rama4], based on a novel model of neuron, complex valued neural networks are designed. Thus, based on the results in section 2, section 3, it is reasoned that transforming real valued signals into complex domain and processing them in the complex domain could have many advantages.

This research paper is organized as follows. In Section 2, Discrete Fourier Transform (DFT) is utilized to transform a set of real/complex valued sequences into the complex valued (frequency) domain. It is reasoned that, in a well defined sense, processing the signals using complex valued neural networks is equivalent to processing them in real domain. In Section 3, a novel model of continuous time neuron is discussed. The associated neural networks (based on the novel model of neuron) are briefly outlined. In Section 4, some important generalizations are discussed. In Section 5, some open questions are outlined. The research paper concludes in Section 6.

II. DISCRETE FOURIER TRANSFORM: SOME COMPLEX VALUED NEURAL NETWORKS:

In the field of Digital Signal Processing (DSP), discrete sequences are processed by discrete time circuits such as digital filters. The transform which converts the time domain information into frequency domain is called as the Discrete Fourier Transform (DFT). One of the main reasons for utilizing the DFT in many applications is the existence of a fast algorithm to compute DFT. This fast algorithm is called as the Fast Fourier Transform (FFT). In the following, we provide the mathematical expressions for the Discrete Fourier Transform (DFT) as well as Inverse Discrte Fourier Transform (IDFT) of a discrete sequence $\{ x_n \}_{n=0}^{M-1}$ i.e. $\{ x_0 , x_1 , x_2 ,......x_{M-1} \}$.

$$\text{DFT:} \quad X(k) = \sum_{n=0}^{M-1} x(n)\, W_M^{k\,n} \quad \text{.....for} \quad 0 \le k \le (M-1) \ldots\ldots\ldots\ldots(2.1)$$

$$\text{IDFT:} \quad x(n) = \frac{1}{M} \sum_{k=0}^{M-1} X(k)\, W_M^{-k\,n} \quad \text{....for} \quad 0 \le n \le (M-1) \ldots\ldots\ldots(2.2)$$

Where

$$W_M = e^{-j(\frac{2\pi}{M})} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(2.3)$$

The results in this section are motivated by the question:

A. ***MAIN QUESTION :*** *Consider a set of samples which are linearly separable in the M-dimensional Euclidean space. Utilizing an invertible (Bijection) Linear Transformation, transform the points. In the transformed domain, are the resulting samples, linearly separable?*

In answering this question, we are led to the following Lemma.

Lemma 1: Under Bijective Linear Transformation, linearly separable patterns in Euclidean Space are mapped to linearly separable patterns in the transform space.

Proof: For the sake of notational convenience, we consider the patterns in a **2**-dimensional Euclidean space. Let the bijective/invertible linear transformation be T: $R^2 \rightarrow R^2$.

Let the original separating line (more generally hyperplane) be given by

$$W_1\, X + W_2\, Y = C \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(2.4)$$

Two regions (decided by the separating line/hyper plane be) in R^2 are:

$$S_1 = \{ \ (x, y) \mid W_1\, x + W_2\, y \ge C \}$$

$$\ldots\ldots\ldots\ldots\ldots..(2.5)$$

$$S_2 = \{ (x, y) \mid W_1\, x + W_2\, y < C \}$$

Now let us consider the Linear Transformation, T:

$$\text{T:} \quad R^2 \rightarrow R^2$$
$$(x, y) \rightarrow (\ p\, x + q\, y\ ,\ r\, x + s\, y\ ) \ldots\ldots\ldots\ldots(2.6)$$

Let the linear transformation be represented by the following matrix:

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.7)$$

Under this transformation, the separating line coordinates become:

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} . \quad \text{……………………………..(2.8)}$$

Thus we readily have

$$\begin{aligned} X' &= p\,X + q\,Y \\ Y' &= r\,X + s\,Y \end{aligned} \quad ,\text{……………………..(2.9)}$$

On inverting the linear transformation, we have

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} p & q \\ r & s \end{pmatrix}^{-1} \begin{pmatrix} X' \\ Y' \end{pmatrix}$$

$$,\text{……………..(2.10)}$$

$$= \begin{pmatrix} s/d & -q/d \\ -r/d & p/d \end{pmatrix} \begin{pmatrix} X' \\ Y' \end{pmatrix}$$

Where d is the determinant of the matrix and is given by $d = p\,s - q\,r$. We thus have

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \frac{s}{d} X' - \frac{q}{d} Y' \\ \frac{-r}{d} X' + \frac{p}{d} Y' \end{pmatrix} \text{…………………….(2.11)}$$

Thus, substituting for X, Y in the original separating line/hyper plane

$W_1\,X + W_2\,Y = C$, we readily have

$$W_1 \left(\frac{s}{d} X' - \frac{q}{d} Y' \right) + W_2 \left(\frac{-r}{d} X' + \frac{p}{d} Y' \right) = C$$

$$\text{…………(2.12)}$$

$$(\,W_1 s - W_2\,r\,)\,X' + (\,-W_1\,q + W_2\,p)\,Y' = C\,d$$

From the above equations, it is clear that a point in two dimensional Euclidean space belonging to the set S_1 gets transformed to the point $T(x.y) = (\,x', \,y'\,)$ i.e

$$\begin{aligned} (x,y) &\in S_1 \\ T(x,y) &= (x', y') \in S_1' \end{aligned}$$

Where the set S_1' is given by

$$S_1' = \{\, (x', y') : (\,W_1 s - W_2\,r\,)\,x' + (\,-W_1\,q + W_2\,p)\,y' \geq C\,d\,\} \text{………..(2.13)}$$

- Thus we have shown that the patterns which are linearly seperable in two dimensional Euclidean space will remain linearly seperable after applying a bijective linear transformation to the samples.
- The above proof is easily generalized to samples in n-dimensional Euclidean space (where 'n' is arbitrary). Q.E.D

Consider the equation (2.1) for computing the Discrete Fourier Transformation of a discrete sequence of samples $\{\, x(n) :\ 0 \leq n \leq (M-1)\,\}$. Let the column vector containing these samples be given by Y. Also, let the column vector containing the transformed samples i.e $\{\, X(k) :\ 0 \leq k \leq (M-1)\,\}$ be given by Z. It is clear that equation (2.1) is equivalent to the following:

Z = F Y,……………………………………..(2.14),

Where F is the Discrete Fourier Transform matrix. This matrix is invertible. Hence the transformation between the discrete sequence vectors Y, Z is bijective. Thus the above Lemma applies.

B. Complex Valued Perceptron:

Consider a single layer of conventional perceptrons. Let the sequence of input vectors be $\{ Y_1, Y_2, ..., Y_L \}$ .The following supervised learning procedure is utilized to classify the patterns:

- Apply the DFT to the successive input training sample vectors resulting in the vectors $\{ Z_1, Z_2, ..., Z_L \}$.
- Train a single layer of Complex Valued Perceptrons using the transformed sample vectors (Complex valued version of Perceptron learning law provided in [AAV] is used)
- Apply the IDFT to arrive at the proper class of training samples.
- Utilize the trained complex valued neural network to classify the test patterns.

In view of Lemma 1, the above procedure converges when the training samples are linearly seperable. Thus the linearly seperable test patterns are properly classified.

The above procedure is also applied for non-linearly seperable patterns using a complex valued Multi-Layer Perceptron. Back propagation algorithm discussed in [Nit1, Nit2] is utilized. Detailed discussion is provided in [Rama1]. *It is argued by Nitta et.al; that the complex valued version of back propagation algorithm converges faster than the real one. Thus from computational view point, the above procedure is attractive.*

III. NOVEL MODEL OF A NEURON: ASSOCIATED NEURAL NETWORKS:

In conventional model of neuron, weighted contribution (weights being the synaptic weights) of *current* input values is taken and a suitable activation function (Signum or Sigmoid or hyperbolic tangent) is applied. A biologically more probable model takes the following facts into account

- The output of a neuron depends *not only* on the current input value, but all the input values over a finite horizon. Thus inputs to neurons are defined over a finite horizon (rather than a single time point).
- Synapses are treated as *distributed elements* rather than lumped elements. Thus synaptic weights are functions defined on a finite support.

For the sake of convenience, let the input as well as synaptic weight functions be defined on the support [0, T].

A. Mathematical Model of Neuron:

Let the synaptic weights be $w_i(t)$, $1 \leq i \leq M$ i.e time functions defined on the support [0,T]. Also, let the inputs be given by $a_i(t), 1 \leq i \leq M$.

Thus, the output of the neuron is given by

$$y(t) = Sign\ (\sum_{j=1}^{M} a_j(t)\ w_j(t) \ ) \text{ ……………..(3.1)}$$

More general activation functions (sigmoid, hyperbolic tangent etc) could be used. The successive input functions are defined over the interval [0,T]. They are fed as inputs to the continuous time neurons at successive SLOTS. For the sake of notational convenience, we call such a neuron, a *continuous time perceptron.*

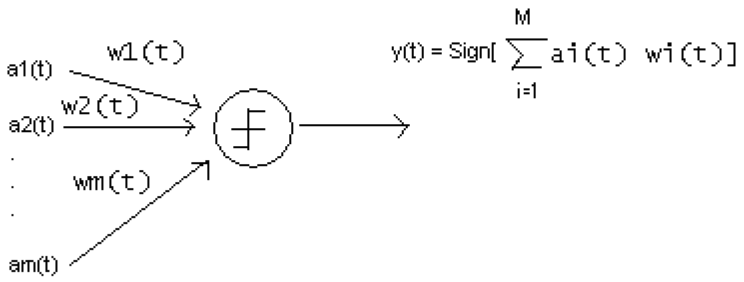

Figure 1. A Novel Model of Continuous Time Neuron.

B. *Continuos Time Perceptron Learning Law:*

As in the case of "conventional perceptron", a continuous time perceptron learning law is given by::

$$W_i^{(n+1)}(t) \;=\; W_i^{(n)}(t) \;+\; \eta\,(\,S(t) - g(t)\,)\,a_i(t)\,, \ldots\ldots\ldots\ldots(3.2)$$

where S(t) is the target output for the current training example, g(t) is the output generated by the continuous time perceptron and η is a positive constant called the learning rate. The proof of convergence of conventional perceptron learning law, also guarantees the point wise convergence (not necessarily uniform convergence) of synaptic weight functions.

Using sigmoid function as the activation function and the continuous perceptron as the model of neuron, it is straightforward to arrive at a continuous time Multi-Layer Perceptron. The conventional back propagation algorithm is generalized to such a feed forward network.

C. *Modulation Theory: Feed Forward Neural Networks:*

Suppose the synaptic weight functions are chosen as sinusoids i.e. $w_i(t) = \cos \upsilon_i t \;\; or \;\; \sin \upsilon_i\, t$ (where $\upsilon_i = 2\,\pi\, f_i$ and f_i's are frequencies of the sinusoids). The weighted contribution at each neuron actually corresponds to Amplitude Modulation (where the synaptic weight functions are the carrier frequencies and the inputs are the base band signals).

We seriously expect that the well known results in Modulation Theory (of communication systems) could be effectively utilized in supervised learning using a single/multiple layer perceptron.

IV. SOME IMPORTANT GENERALIZATIONS:

- Unlike the perceptron model (inputs constitute a vector) discussed previously, it is possible to consider the case where the inputs constitute a three/multi-dimensional array (For instance in biological systems, the neurons are indexed by three dimension variables). Utilizing Tensor products, the outputs of continuous time neurons are obtained. Also, using the above model of neuron, multi-layer, multi-dimensional neural networks (such as Multi-dimensional Mult-Layer Perceptron) are designed and studied [Rama1].
- Based on the above model of neuron, it is possible to consider complex valued neural networks in which the input functions, synaptic weight functions, thresholds are complex valued. It is possible to generalize the perceptron learning law, complex valued back propagation algorithm to such complex valued neural networks [Rama1].
- It should be possible to design and study complex valued associative memories based on the above model of neuron.

V. SOME OPEN QUESTIONS:

- Is it possible to generalize Lemma 1 (discussed in section 2) to function spaces? Or equivalently, what is the most general version of Lemma 1 ?
- Consider the problem of supervised learning in a function space. Equivalently consider a function space over [0,T]. Define a distance metric over such a space. Design a neural network which can be trained to classify the patterns into finitely many classes (of functions) [Rama4].
- CLIFFORD NEURAL NETWORKS: Some researchers modeled the neuronal parameters using quaternions. These quaternion based neural networks are utilized in practical applications such as colour night vision [KIM]. Also some authors have utilized geometric algebra in designing novel neural networks. An important open problem is to show that the Clifford/geometric algebra based neural networks have important advantages over the real valued neural networks.

VI. CONCLUSIONS:

In this paper, transforming real valued signals into complex domain (using DFT) and processing them using complex valued neural network is discussed. A novel model of neuron is proposed. Based on such a model real as well as complex valued neural networks are proposed. Some open research questions are provided.

REFERENCES:

[AAV] I. N. Aizenberg, N. N. Aizenberg and J. Vandewalle, “Multi-Valued and Universal Binary Neurons,” Kluwer Academic Publishers, 2000.

[Hir] A.Hirose, “Complex Valued Neural Networks: Theories and Applications, “World Scientific Publishing Company, November 2003.

[KIM] H.Kusamichi, T.Isokawa, N.Matsui et.al, “A new scheme for colour night vision by quaternion neural network, “ 2nd International Conference on Autonomous robots & agents, Dec. 13-15, 2004, Palmerston North, New Zealand

[Nit1] T. Nitta and T. Furuya: "A Complex Back-propagation Learning", Transactions of Information Processing Society of Japan, Vol.32, No.10, pp.1319-1329 (1991) (in Japanese).

[Nit2] T. Nitta : "An Extension of the Back-Propagation Algorithm to Complex Numbers", Neural Networks, Vol.10, No.8, pp.1391-1415 (1997).

[Rama1] G. Rama Murthy, “Multi-Dimensional Neural Networks: Unified Theory of Control, Communication and Computation, “ Research Monograph considered by Pearson Education Publishers, New York

[Rama2] G. Rama Murthy, “Multi/Infinite Dimensional Neural Networks, Multi/Infinite Dimensional Logic Theory, “ International Journal of Neural Systems, Vol.15, No. 3 (2005), 1-13, June 2005

[Rama3] G. Rama Murthy and D. Praveen, “Complex-Valued Neural Associative Memory on the Complex Hypercube, “ Proceedings of 2004 IEEE International Conference on Cybernetics and Intelligent Systems (CIS-2004), Singapore..

[Rama4] G. Rama Murthy, “Linear filter model of synapses: Associated novel real/complex valued neural networks, “ Accepted for Publication in the Proceedings of 9th Fuzzy Days (International Conference on Computational Intelligence, September 2006, Dortmund, Germany.