27 Introduction

(1982) J. J. Hopfield

Neural networks and physical systems with emergent collective computational abilities

Proceedings of the National Academy of Sciences 79:2554–2558

As far as public visibility goes, the modern era in neural networks dates from the publication of this paper by John Hopfield. This paper is short, clearly written, and brings together a number of strands that up to this point were somewhat separated in the neural network literature. Much of the reason for its impact was simply this: it presented a sophisticated, coherent theoretical picture of how a neural network could work, and what it could do.

We cannot avoid making a comment about the sociology of science, it is hoped without causing offense. John Hopfield is a distinguished physicist. When he talks, people listen. Theory in his hands becomes respectable. Neural networks became instantly legitimate, whereas before, most developments in networks had been the province of somewhat suspect psychologists and neurobiologists, or by those removed from the hot centers of scientific activity.

Other well known physicists represented in this volume had also done distinguished work in neural networks: Leon Cooper, Gordon Shaw, and William Little. Although their work was noted and respected, it was subliminal, as far as the scientific world at large was concerned. The models they proposed were brain models first and useful devices a distant second. Practical implications, though clearly present, were not emphasized.

The one thing that really seems to have made the Hopfield work take fire in terms of public notice was the immediate and strong contact he made with the new chip building technology that was finally capable of constructing the devices he was proposing. The first attempts to make chips followed within a couple of years of this 1982 paper, and by early 1987, AT&T Bell Laboratories had announced successful development of neural net chips, largely based on the Hopfield networks (*Electronics*, March 5, 1987, p. 21) and Carver Mead and coworkers (paper 43) were making artificial sensory systems using VLSI technology, inspired by initial contact with Hopfield. The Caltech environment and the potential usefulness of the neural networks Hopfield discussed made the engineering connection immediate.

The criticism has been made by some old-timers in the neural network field that there was nothing fundamentally new in the model proposed by Hopfield. We have collected in this volume a number of earlier papers, and can let readers draw their own conclusions on this point. Although many of the ideas in this paper have precursors, as Hopfield would be the first to admit (see his list of references!), bringing them all together, with detailed, clear, and powerful mathematical analysis, is creative work of the first order, and the paper richly merits the attention and respect it has received.

There are a number of technical points that are worth mentioning. The order of presentation of ideas that Hopfield uses is the opposite of that used by most network modelers. The standard approach to a neural network is to propose a learning rule, usually based on synaptic modification, and then to show that a number of interesting effects arise from it. Hopfield starts by saying that the function of the nervous system is to develop a number of locally stable points in state space. Other points in state space flow into the stable points (called attractors). This allows a mechanism for correcting errors, since deviations from the stable points disappear. It can also reconstruct missing information since the stable point will appropriately complete missing parts of an incomplete initial state vector.

Hopfield then proceeds to develop a network that shows this desired behavior. He assumes that the basic elements of the network are threshold logic units, which sum synaptic inputs, compare the sum with a threshold, and then respond 1 if the sum is at or above threshold and 0 otherwise. In a later paper (paper 35) Hopfield discusses networks of neurons that can show graded intermediate states. The network is recurrent, in that the neurons connect to each other, with the exception that a neuron does not connect to itself; that is, the connection matrix has zeros down the main diagonal.

Hopfield assumes that the system wants to learn a set of states, $\{V^s\}$, with individual element activities V_i. He suggests a learning rule for constructing elements of the connectivity matrix, which is the Hebb rule, combined with scaling terms, for placing the point for zero connection modification at an activity of one-half. This ensures symmetry of modification magnitude for the two allowable output states of a cell, 0 and 1.

In one paragraph Hopfield suggests one of the most important new techniques to have been proposed in neural networks. He considers the special case of a symmetric matrix, i.e., ones where $\mathbf{T}_{ij} = \mathbf{T}_{ji}$. Then he defines a quantity, called E, which is the sum of all the terms:

$$E = -\tfrac{1}{2} \sum_{i \neq j} \sum \mathbf{T}_{ij} V_i V_j.$$

This term is equivalent to physical *energy*. As the system evolves, due to the feedback dynamics, the energy decreases until it reaches a (perhaps local) minimum. Hopfield next makes the portentous comment, “This case is isomorphic with an Ising model,” thereby allowing a deluge of physical theory (and physicists) to enter network modeling. This flood of new participants has transformed the field of neural networks.

The dynamics of evolution of the system state follows a simple rule and is asynchronous. An element, chosen at random, looks at its inputs, and changes state, depending on whether or not the sum of its input is above or below threshold. It can be seen from the form of the energy term that a state change leads either to a decrease in energy or to the energy remaining the same. The updating rule is, therefore, an energy minimizing rule. Modifications of element activities continue until a stable state is reached, that is, a energy minimum is reached.

A number of computer simulations and some analysis led Hopfield to conclude that the number of ‘memories’ that could be stored accurately by a network was about 15% of the dimensionality. This number agrees well with experience of others. It has also led to a number of attempts to increase storage capacity by various techniques, as well as some more accurate definitions of and computations with storage capacity. However, *most* reasonable models, with *most* reasonable definitions of capacity, end up having a capacity of 10–20% of the number of elements. A number of estimates of capacity can be found in a volume of papers growing out of the 1986 Neural Networks for Computation Conference (Denker, 1986).

Reference

J. S. Denker (Ed.) (1986), *Neural Networks for Computation*, AIP Conference Proceedings 151, New York: American Institute of Physics.