Neural Network Architectures

An Introduction

Judith E. Dayhoff

VNR VAN NOSTRAND REINHOLD
New York

To the memory of Margaret Oakley Dayhoff, pioneer in evolutionary biology.

Library of Congress Catalog Number 89-78147
ISBN 0-442-20744-1

Printed in the United States of America

Van Nostrand Reinhold
115 Fifth Avenue
New York, New York 10003

Van Nostrand Reinhold International Company Limited
11 New Fetter Lane
London EC4P 4EE, England

Van Nostrand Reinhold
480 La Trobe Street
Melbourne, Victoria 3000, Australia

Nelson Canada
1120 Birchmount Road
Scarborough, Ontario M1K 5G4, Canada

16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Dayhoff, Judith E.
Neural network architectures : an introduction / by Judith E. Dayhoff.
p. cm.
ISBN 0-442-20744-1
1. Computer architecture. 2. Neural computers. 3. Computer networks. I. Title.
QA76.9.A73D39 1990
006.3—dc20 89-78147
CIP

Neural Network Architectures

Preface

This book was written as a means of introducing both artificial and biological neural networks to a disparate audience comprising biologists and engineers as well as business professionals and others. More than an introduction, this text makes an excellent compendium for those who have already begun to experiment with neural networks. This is done by focusing on neural network paradigms, which are of equal use to beginners and professionals. Neural network paradigms are defined, studied, and explained by providing example applications and the equations that govern each network's computations.

This book also covers biological neural systems in a way that is clear and readable to non-specialists. Considerable depth is provided in the areas of biological neurons and their synapses, as these are the entities that we hope some day to emulate. The reader will be exposed to the remarkable complexity of biological neural systems as compared to artificial neural networks.

Emphasis is given to applications and examples using neural networks. An entire chapter is dedicated to backpropagation applications (Chapter 5), and another gives a general, comprehensive description of the applications of neural nets (Chapter 11). Each chapter that covers a neural network paradigm also includes applications for that paradigm.

The reader, after completing this book, will be able to understand the basic neural network paradigms and will grasp the general ideas behind neural network design. He or she will learn the steps of using an artificial neural network, will become familiar with a broad range of applications possibilities, and will be ready to begin designing experiments.

ACKNOWLEDGMENTS

I would like to thank those who helped to bring this book into existence. Appreciation goes to the readers of the manuscript: Edward Dayhoff, Dale Mugler, Wilson Or, Steve Fuld, Craig Will, Bruce Lindsey, Margaret Mortz, and Nancy Dayhoff, for their valuable feedback, insights, and encouragement, and to the artist, Phil Monti, who once again has proven that a picture is worth a thousand words.

Neural Network Architectures

Contents

1
Introduction

Neural networks provide a unique computing architecture whose potential has only begun to be tapped. Used to address problems that are intractable or cumbersome with traditional methods, these new computing architectures — inspired by the structure of the brain — are radically different from the computers that are widely used today. Neural networks are massively parallel systems that rely on dense arrangements of interconnections and surprisingly simple processors.

Artificial neural networks take their name from the networks of nerve cells in the brain. Although a great deal of biological detail is eliminated in these computing models, the artificial neural networks retain enough of the structure observed in the brain to provide insight into how biological neural processing may work. Thus these models contribute to a paramount scientific challenge — the brain understanding itself.

Neural networks provide an effective approach for a broad spectrum of applications. Neural networks excel at problems involving patterns — pattern mapping, pattern completion, and pattern classification. Neural networks may be applied to translate images into keywords, translate financial data into financial predictions, or map visual images to robotic commands. Noisy patterns — those with segments missing — may be completed with a neural network that has been trained to recall the completed patterns (for example, a neural network might input the outline of a vehicle that has been partially obscured, and produce an outline of the complete vehicle).

Possible applications for pattern classification abound: Visual images need to be classified during industrial inspections; medical images, such as magnified blood cells, need to be classified for diagnostic tests; sonar images may be input to a neural network for classification; speech recognition requires classification and identification of words and sequences of words. Even diagnostic problems, where results of tests and answers to questions are classified into appropriate diagnoses, are promising areas for neural networks. The process of building a successful neural network application is complex, but the range of possible applications is impressively broad.

Neural networks utilize a parallel processing structure that has large numbers of processors and many interconnections between them. These processors are much simpler than typical central processing units (CPUs). In a neural network each processor is linked to many of its neighbors (typically hundreds or thousands) so that there are many more interconnects than processors. The power of the neural network lies in the tremendous number of interconnections.

Neural networks are generating much interest among engineers and scientists. Artificial neural network models contribute to our understanding of biological models, provide a novel type of parallel processing that has powerful capabilities and potential for creative hardware implementations, meet the demand for fast computing hardware, and provide the potential for solving applications problems.

Neural networks excite our imagination and relentless desire to understand the self, and in addition equip us with an assemblage of unique technological tools. But what has triggered the most interest in neural networks is that models similar to biological nervous systems can actually be made to do useful computations, and, furthermore, the capabilities of the resulting systems provide an effective approach to previously unsolved problems.

In this volume we introduce a variety of different neural network architectures, illustrate their major components, and show the basic differences between neural networks and more traditional computers. Ours is a descriptive approach to neural network models and applications. Included are chapters on biological systems that describe living nerve cells, synapses, and neural assemblies. The chapters on artificial neural networks cover a broad range of architectures and example problems, many of which can be developed further to provide possibilities for realistic applications.

TRADITIONAL VERSUS NEURAL NETWORK ARCHITECTURE

Neural network architectures are strikingly different from traditional single-processor computers. Traditional Von Neumann machines have a single CPU that performs all of its computations in sequence. A typical CPU is capable of a hundred or more basic commands, including adds, subtracts, loads, and shifts, among others. The commands are executed one at a time, at successive steps of a time clock. In contrast, a neural network processing unit may do only one or, at most, a few calculations. A summation function is performed on its inputs; incremental changes are made to parameters associated with interconnections. This simple structure nevertheless provides a neural network with the capabilities to classify and recognize patterns, to perform pattern mapping, and to be useful as a computing tool.

The processing power of a neural network is measured mainly by the number of interconnection updates per second; in contrast, Von Neumann machines are benchmarked by the number of instructions that are performed per second, in sequence, by a single processor. Neural networks, during their learning phase, adjust parameters associated with the interconnections between neurons. Thus, the rate of learning is dependent on the rate of interconnection updates.

Neural network architectures depart from typical parallel processing architectures in some basic respects. First, the processors in a neural network are massively interconnected. As a result, there are more interconnections than there are processing units. In fact, the number of interconnections usually far exceeds the number of processing units. State-of-the-art parallel processing architectures typically have a smaller ratio of interconnections to processing units. In addition, parallel processing architectures tend to incorporate processing units that are comparable in complexity to those of Von Neumann machines. Neural network architectures depart from this organization scheme by containing simpler processing units, which are designed for summation of many inputs and adjustment of interconnection parameters.

BIOLOGICAL NEURAL SYSTEMS — THE ORIGINAL NEURAL NETWORKS

Neural network architectures are motivated by models of our own brains and nerve cells. Although our current knowledge of the brain is limited, we do have much detailed anatomical and physiological information. The basic anatomy of an individual nerve cell — or neuron — is known, and the most important biochemical reactions that govern its activities have been identified.

A diagram of a nerve cell typical of those in the human brain is shown in Figure 1-1. The output area of the neuron is a long, branching fiber called the axon. An impulse can be triggered by the cell, and sent along the axon branches to the ends of the fibers. The input area of the nerve cell is a set of branching fibers called dendrites. The connecting point between an axon and a dendrite is the synapse. When a series of impulses is received at the dendritic

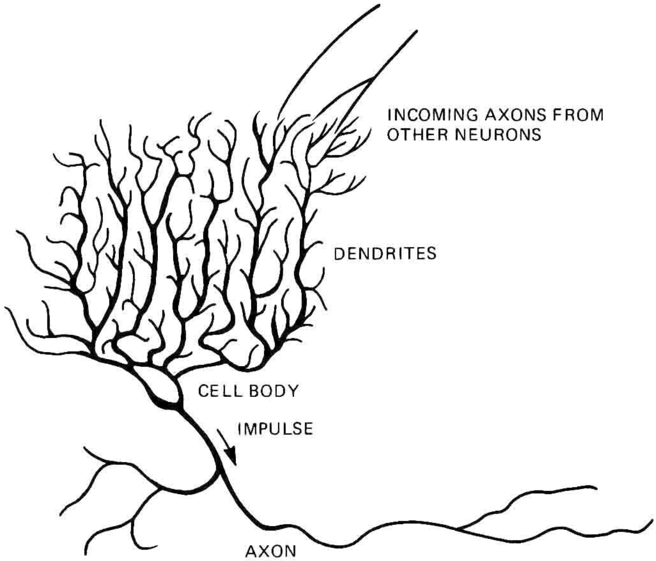

Figure 1-1. Schematic drawing of a biological nerve cell.

areas of a neuron, the result is usually an increased probability that the target neuron will fire an impulse down its axon.

The neuron shown in Figure 1-2a was photographed from a tissue culture of embryonic nerve cells. Although the axon is hidden, the dendritic tree is

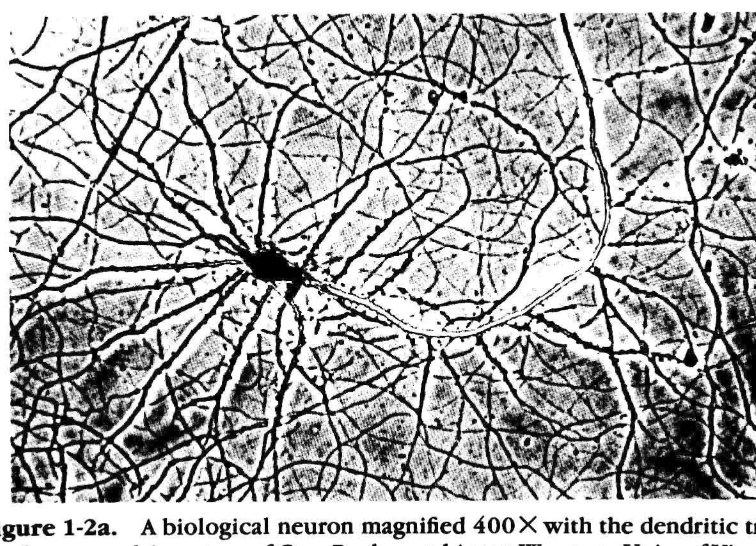

Figure 1-2a. A biological neuron magnified 400× with the dendritic tree in the foreground (courtesy of Gary Banker and Aaron Waxman, Univ. of Virginia).

apparent. The many larger fibers in the foreground are dendritic branches; the smaller fibers that crisscross in the background are axons that synapse onto the dendrites, bringing incoming impulses from other neurons.

Figure 1-2b shows a typical network of neurons, traced from the human visual cortex. These neurons appeared when a thin section of the cortex was impregnated with a Golgi stain, which is taken up by only 2% of the neurons.

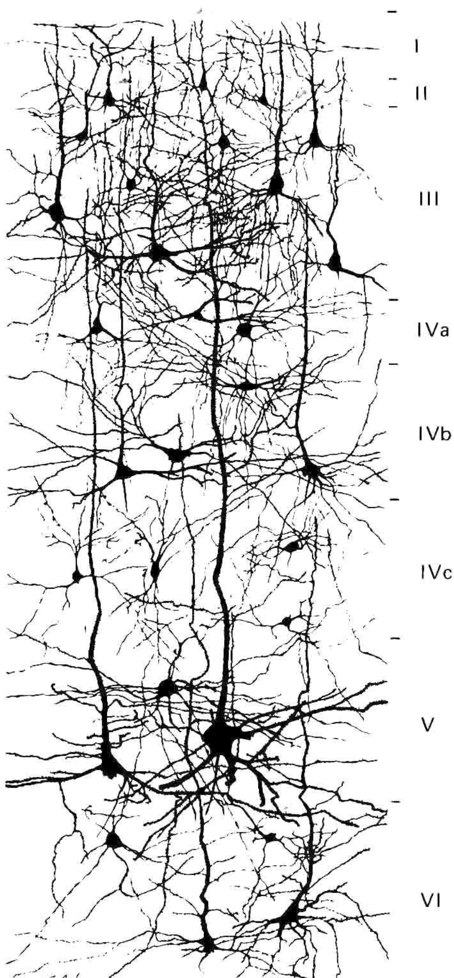

Figure 1-2b. A Golgi-stained preparation from the visual cortex of a two-year-old child showing prominent dendritic arborizations (from Conel, *The Postnatal Development of the Human Cerebral Cortex,* vol VI. Harvard Univ. Press, 1959).

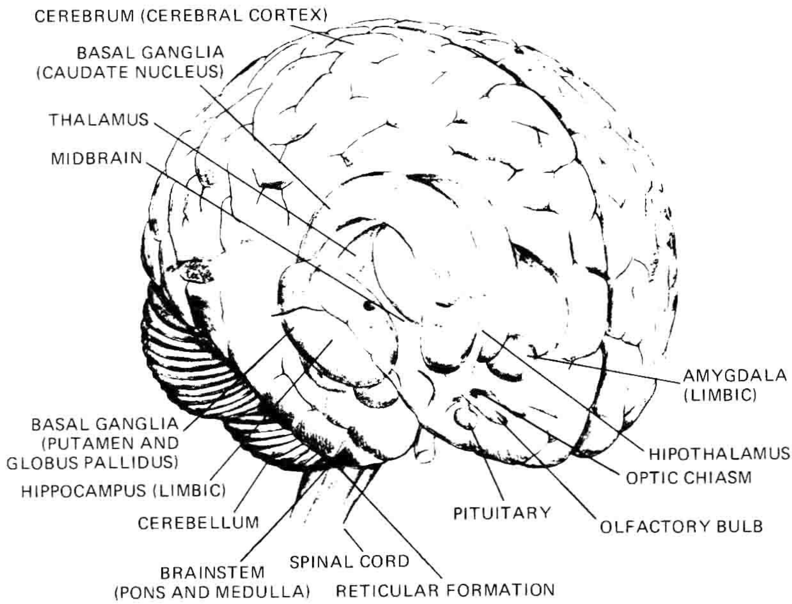

Figure 1-3. Major structures of the human brain (from Nauta and Feirtag, The organization of the brain, *Scientific American* 1979).

The resulting picture indicates the nature of the biological neural network present, with densely placed neurons and myriad intersecting nerve branches. (The actual biological network is much more dense than that shown in the figure because of the sparsity of cells that take up the Golgi stain.) This picture exemplifies the vast interconnected arrays of neurons that appear in biological neural networks.

Figure 1-3 depicts the human brain. The basic circuitry of the brain is considered in terms of general pathways. Details concerning which individual neurons are connected to which other individual neurons have not yet been mapped in the human nervous system, but considerable research effort has been put toward elucidating the detailed circuitry of the brain and determining both the fixed structure and the degree of flexibility present.

The brain is a dense neural network in which the neurons are highly interconnected. The total number of neurons in the human brain is estimated at 100 billion (DARPA Neural Network Study, 1988). Each neuron is connected to perhaps 10,000 other cells, meaning each biological neuron can send impulses that may be received by as many as 10,000 target cells.

Figure 1-4 shows a comparison of different biological nervous systems with artificial neural networks (DARPA, 1988). Speed, in terms of interconnections processed per second, is plotted against storage, measured in terms of

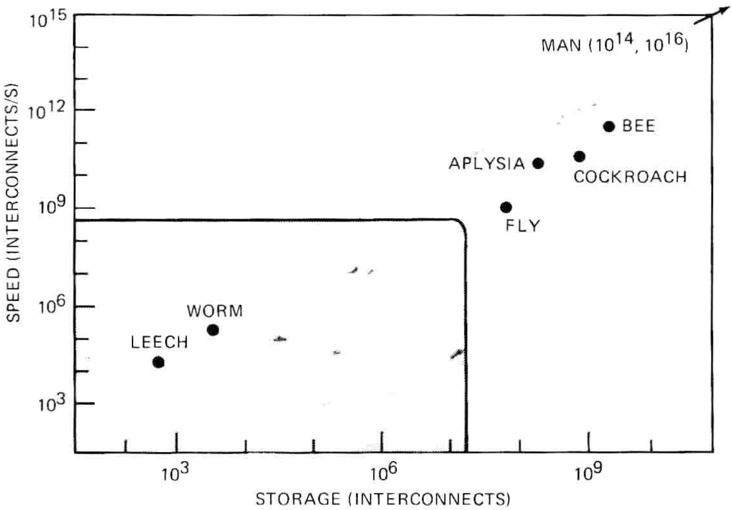

Figure 1-4. Speed versus storage for a variety of systems. Speed is measured in interconnects per second (vertical axis) and storage is measured in interconnects (horizontal axis). The shaded area shows the power of existing simulators (from DARPA Neural Network Study, 1988).

interconnections. The shaded area represents neural network sizes that are within the reach of today's artificial neural net simulations. The leech and worm, relatively primitive invertebrates, have nervous systems that appear within the range of existing simulators having fewer than 10^8 interconnections. More complex organisms, such as the fly, bee, cockroach, and aplysia (a sea slug), have nervous systems with considerably more speed and storage capacity. They appear to exceed the computational capabilities presently available in simulations. The human nervous system is far larger than the other systems plotted, and would appear beyond the top right of the graph.

ARTIFICIAL NEURAL NETWORKS—THE BASIC STRUCTURE

Figure 1-5 depicts an example of a typical processing unit for an artificial neural network. On the left are the multiple inputs to the processing unit, each arriving from another unit, which is connected to the unit shown at the center. Each interconnection has an associated connection strength, given as w_1, w_2, . . ., w_n. The processing unit performs a weighted sum on the inputs and uses a nonlinear threshold function, f, to compute its output. The calculated result is sent along the output connections to the target cells

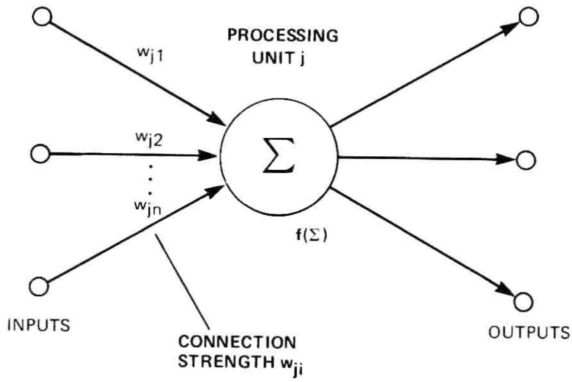

Figure 1-5. Schematic processing unit from an artificial neural network.

shown at the right. The same output value is sent along all the output connections.

The neural network shown in Fig. 1-6a has three layers of processing units, a typical organization for the neural net paradigm known as back-error propagation. First is a layer of input units. These units assume the values of a pattern, represented as a vector, that is input to the network. The middle, "hidden," layer of this network consists of "feature detectors"—units that respond to particular features that may appear in the input pattern. Sometimes there is more than one hidden layer. The last layer is the output layer. The activities of these units are read as the output of the network. In some applications, output units stand for different classifications of patterns.

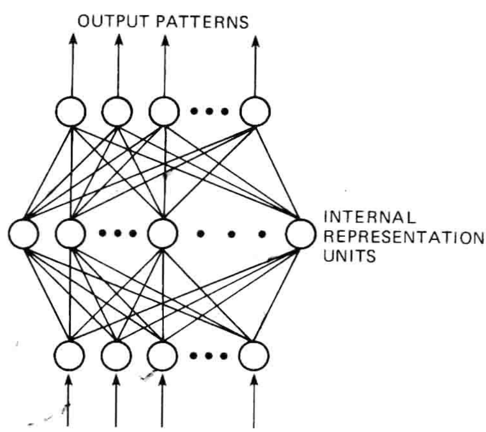

Figure 1-6a. An artificial neural network with three fully interconnected layers (from Rumelhart and McClelland, *Parallel Distributed Processing*. MIT Press, 1986).

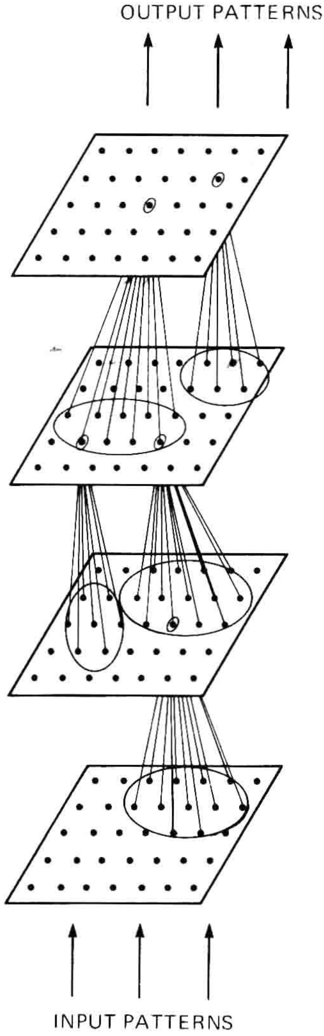

Figure 1-6b. A multilayered network with slabs of processing units that are interconnected with adjacent layers (from DARPA Neural Network Study, 1988).

A larger neural network, in which each layer is organized as a two-dimensional slab of neurons, is shown in Figure 1-6b. Neural networks are not limited to three layers, and may utilize a huge number of interconnections.

Each interconnection between processing units acts as a communication route: Numeric values are passed along these interconnections from one