SLAC-PUB-5188
February 1990
(M)

Neural Networks: Capabilities & Applications †

ROBERT C. SASS

Stanford Linear Accelerator Center
Stanford University, Stanford, California 94309

ABSTRACT

Make some generalizations about the capabilities of neural networks and identify the areas where they do well and where they don't. Survey a variety of applications where neural networks have been used. Define general criteria for identifying good neural network applications and propose a longer term plan.

Submitted for Publications

† Work supported by the Department of Energy, Contract DE–AC03–76SF00515.

TABLE OF CONTENTS

1 Introduction 3

2 Why Neural Networks? 4

3 A Critical Assessment of Neural Network Technology 5

3.1 Pattern Classification 6

3.2 Optimization 6

3.3 Self-Organization/Category Formation 7

3.4 Associative Memory 7

4 Applications 8

4.1 Process Monitor 9

4.2 Risk Analysis 9

4.3 Fork Lift Robot 10

4.4 Explosive Detection 10

4.5 Broom Balancer 11

4.6 Sonar Classifier 11

4.7 Miscellaneous 11

5 Application Selection Criteria 12

6 A Modest Proposal 13

REFERENCES 15

1. Introduction

In recent years there has been a great resurgence of interest in a method of computing that was originally pioneered in the 1940s. This method is modeled generally after biological nervous systems† (wetware) and is called neural networks (NN), artificial neural networks (ANN), parallel distributed processing (PDP) and perhaps others. In view of the large amount of research and application activity going on in this area, it seems appropriate to make a high level overview of the field to see if this approach can be usefully applied to computing applications at SLAC.

Much of the information presented in this paper was gleaned from the DARPA Neural Network Study[1] which was published in November of 1988. The study was prompted as part of a Balanced Technology Initiative for automatic target recognition problems and smart weapons and eventually developed into a national study on neural networks. -The study director was Prof. Bernard Widrow of Stanford University, with support provided by the MIT/Lincoln Laboratory staff and scientists. During the five-month course of study, "representatives of every school of thought in neural networks were consulted."

This paper does not review details of neural network models in general or any particular model; there are a number of good references for that.[2][3] It is instead an attempt to provide an overview of the current state of neural network capabilities and applications; where they are good and where they are not. It also provides a list of criteria for their application so you can begin thinking about their utility in your particular area of interest.

In a real sense, neural networks are still a solution in search of a few good problems. I think you'll see from the broad range of applications to which they

† The often-quoted numbers are that the human cortex alone contains about 10^{11} neurons, each with about 10^3 connections giving a total of 10^{14} connections. Since neurons operate at about 100 Hz, the human cortex processes about 10^{16} interconnects/second, far beyond any present computer's capabilities.

are being applied, there are probably many good problems areas at SLAC and the SLC where a neural network approach could be useful. Some general application areas include beam steering, tuning and diagnostics, error analysis, CID bunching, positron production, etc.

2. Why Neural Networks?

In the most general terms, a neural network is: "a system composed of many simple interconnected processing elements operating in parallel whose function is determined by network structure, connection strengths and the computation performed at the processing elements." The revival of interest in this approach to computing is partly due to the failure of conventional programming techniques, including rule-based AI, to solve some of the "hard" problems like machine vision, continuous speech recognition and machine learning. A neural network approach has some inherent capabilities which other programming techniques lack.

1. They are naturally parallel and thus hold the promise of being able to solve "hard" problems which require very large amounts of conventional, serial computational power. As suitable hardware implementations are developed over the next few years,† it is expected that they will be increasingly applied to these problems which must be computed in real-time or at least reasonable time.

2. They learn either in a supervised mode, where the network is provided with the correct response, or in an unsupervised mode, where the network self-organizes and extracts patterns from the data presented to it. After training, the network generalizes to produce a correct response to input combinations for which it has not been trained.

3. A three-layer network can perform any continuous non-linear mapping from inputs to outputs or emulate any deterministic classifier, and thus for some

† A spinoff from DEC called MASPAR has recently announced a massively parallel machine with up to 16K processors that will be priced in the range of a high end VAX.

problems they are easier to implement than conventional pattern matching or statistical methods.

4. There have been significant demonstrations of neural network capabilities in vision, speech, signal processing, and robotics. The variety of problems addressed by neural networks is impressive.

In the midst of all the press and activity in the field, we should consider this quote from Carver Mead who observed, "In a field like neural networks, one is usually too optimistic in the short run, but one is never optimistic enough over the long run."

3. A Critical Assessment of Neural Network Technology

There are about 30 different neural network models which have been developed so far. The motivations for these models run the gamut from the emulation and study of biological systems to the explicit solution of optimization problems. However, the kinds of tasks that all of them perform can be broken down into a few general but overlapping catagories. In the following sections, we'll look at these categories and see how neural networks compare with existing technologies.

It should be kept in mind that much of the criticism which follows reflects neural network architectures simulated on conventional machines. Performance problems with some neural network implementations may disappear with the advent of large-scale parallel architectures. Scaling problems, which at first glance seem to represent a more fundamental limitation that could limit the utility of a particular set of models, can often be overcome in the same way they were overcome in the construction of large conventional software systems: modularization. Multiple, smaller, interacting networks can often achieve the same result while avoiding the scaling problems. As you might expect, in some applications neural networks are equal (or more rarely, superior) to standard approaches either in performance or ease of implementation. In others, they are clearly not competitive at this time.

3.1 PATTERN CLASSIFICATION

Trainable pattern classifiers may be one of the best practical uses for neural networks. They do about as well as standard statistical pattern classifiers (nearest neighbor) and are easier to implement. The network is trained with known data and learns to classify the input into a pre-specified number of catagories. The inputs and outputs are dependent on the specific problem and can be binary or continuous. There are a wide range of "interesting" problems which can be cast in terms of pattern recognition or classification. The patterns classified can represent words for a speech recognizer, objects for a visual image classifier, decisions in analysis applications, outputs in a control system, etc. Some implementations in this category are sonar and radar target classification, signal processing, Kanji and written character recognition, robot control, conventional process control,[4][5] risk evaluation on loan applications, text-to-speech translation, etc. Again, many problems can be represented as complex, non-linear pattern classifiers.

Since these neural network models executing on a serial computer compare favorably with conventional algorithms, the advent of VLSI neural chips should offer dramatic gains in this area. Additionally, because the network learns the pattern classification, its implementation is more straightforward than the modeling and design of a fixed conventional algorithm.

3.2 OPTIMIZATION

The Traveling Salesman is representative of a category of optimization problems where the solution space rises exponentially with the size of the problem. It can be briefly stated as follows: given a list of cities and the distances between each pair, find the minimum length tour that visits each city just once. The Hopfield Net [6] restates the problem as one of energy minimization and uses a fully connected neural network to solve it for a small number (10) of cities. While theoretically interesting, the Hopfield Net doesn't optimize nearly as well as other conventional algorithms. Furthermore, it has severe scaling problems and only works at all with small problem spaces. By way of comparison, the Hopfield Net with 30 cities only gets within 17% of optimum. Conventional optimization algorithms get within 2% and scale to 50,000 cities.

A simulated annealing algorithm can handle much larger problem spaces and is very general, but is computationally slower than standard algorithms applied to specific problems. Whether the simulated annealing network can scale to that of standard algorithms is unclear. The performance difference could be compensated for with a parallel architecture, but these are not yet readily available.

We conclude that solving optimization problems with neural networks is of theoretical interest and its implementation may be easier than a conventional approach. However, the selection of a neural network to solve a specific optimization is valid only if the problem domain is relatively small and doing the analysis to apply the appropriate conventional algorithm is not justified.

3.3 SELF-ORGANIZATION/CATEGORY FORMATION

These models perform pattern classification where the categories are not known *a priori*. This type of clustering is useful for information reduction where large amounts of uncategorized training data are available. They are very popular for the exploration of the biological basis of learning and behavior as well as speech and vision applications. Since speech and vision constitute large areas in themselves, these models are of real interest and value. Their role in traditional computing applications and data analysis is less clear, but as hardware becomes available, they can be used for pre-processing large amounts of data in real-time.

3.4 ASSOCIATIVE MEMORY

In these models, the complete memory item is provided from a partial or corrupted input key. They have been extensively studied but real-world applications that use this capability of neural networks are hard to find and interest in them seems to be primarily theoretical.

This section in the DARPA study was a scathing critique of associative memories in general and neural network applications in particular. A few quotes will give the flavor. "The long-term applications of associative memories, as pure memories, have yet to be defined in any detail. Given that they have been studied since the early 1950s with no significant practical application having yet materialized, their importance may be limited."

Looking at potential applications for neural networks as problems in associative memory does not seem productive at this time.

4. Applications

There are a wide variety of applications to which neural networks are being applied, sometimes even successfully! Researchers usually try to apply them to problems that are either difficult or impossible to solve with "classical" approaches. They seem advantageous for problems that are data-intensive, deal with changing or ill-defined environments or for which a model is difficult to construct. Speed, parallelism and ease of implementation are most often cited as reasons for using a neural network approach. Self-organization, unsupervised learning, generalization, fault-tolerance and noise immunity are further advantages not typically found in "classical" approaches.

Most researchers feel that neural networks will enhance rather than supplant more traditional approaches. All of the presently available commercial neural network development systems run on standard PCs or workstations (though some require coprocessors for reasonable performance) and several have the capability to generate code for embedded use in existing applications. At least the field doesn't seem to be repeating the disaster of AI by building special purpose machines that only operate in their own environment.

While the field is very diverse, encompassing both "real-world" engineering problems, anthropomorphically-inspired problems relating to behavior and cognition, and theoretical efforts aimed at understanding and describing the properties of large networks of distributed processors, our focus here will be on those application of more practical interest and which have achieved some degree of success at the present time. Machine vision and continuous speech recognition are areas in which neural networks are being intensively applied and about which there is a great deal of literature. However, because these are "hard" research problems, they are ignored in what follows.

In the DARPA study, 55 researchers reported on 77 applications and of these, 11 were described in some detail. In addition, there are several interesting applications gleaned from trade journals where neural networks have been embedded in products to enhance their overall performance.

4.1 PROCESS MONITOR

In this application, a modified ADELINE network predicts the yield and other performance measures of a fluorescent bulb manufacturing plant from 100-200 sensor measurements. The-function performed by the existing network is the same as that performed by conventional statistics (linear regression) but with significant computational advantages in that the network processes the data incrementally every five minutes whereas conventional regression techniques require data to be collected for a longer period and then processed in batch. Future plans involve using additional networks and moving from process monitoring to process control. There is also an ongoing effort to tie neural networks to the large existing body of control theory.

4.2 RISK ANALYSIS

This is a fielded application and was built using the commercial development system sold by Nestor, Inc. Multiple interacting networks were trained with existing data to be a Mortgage Origination Underwriter, a Mortgage Insurance Underwriter and a Delinquency Risk Processor. A very interesting aspect of this application is that the Nestor system provides the reasons an application was denied based on an analysis of the internal mechanisms employed by the networks.

4.3 FORK LIFT ROBOT

Because neural networks can handle arbitrary non-linearities, there seems to be great potential for their application to some hard control system problems. The DARPA study discusses the CMAC (pronounced "See MAC" here for obvious reasons!) model and its use in the control of a robotic fork lift as opposed to that of standard control methodology which builds a model of the system.

A conventional robot was taught by a human operator to guide the insertion of a forklift into a pallet that is offset in all six dimensions. This is very difficult for conventional control systems because the relationship between sensor patterns and appropriate movement is complicated and nonlinear. Thus, the system learns an arbitrary mapping function from the present and desired states of the manipulator to the actuator control. The neural network learned to acquire the pallet from an -arbitrary starting position with as few as seven teaching points. "The ability to learn complex control functions and to generalize in unexpected situations will allow robots with neural network controllers to function in environments that were previously impossible for traditional control systems." It appears that neural network implementations will surely appear in future robotic systems.

4.4 EXPLOSIVE DETECTION

This is a fielded application in which an embedded neural network replaced conventional pattern-recognition techniques to interpret signals from a thermal neutron analysis device which is part of an explosive detection system being used for airport security. The addition of the network both reduced the false alarms and raised the probability of detection. The device which contains the network met the required FAA specifications and is being sold commercially.

4.5 BROOM BALANCER

In this application, an ADELINE[7] network was trained to balance an inverted pendulum positioned on a cart which could move in two dimensions. Network input was two 5 x 7 pixel images of the present and previous position of the pendulum. From these two images, the network derived the cart and pendulum velocities and positions, and after training was able to balance the inverted pendulum indefinitely.

The inverted pendulum is a classical control problem that has been extensively studied and is representative of many other control problems. This implementation demonstrates that an adaptive network can extract the critical state information of the pendulum and cart's position and velocity from a digitized visual image.

4.6 SONAR CLASSIFIER

This application used the standard back error propagation network to identify active sonar returns as a mine or a rock. The network outperformed both the nearest neighbor (statistical) and human classifiers on the same set of signals. The input to the network was the power spectral envelope of active sonar returns collected from a mine and a rock on a sandy ocean floor. The output of the network was only two units, one indicating "rock" and the other "mine."

4.7 MISCELLANEOUS

The Nestor Learning System, a commercial neural network development system, has been applied to a number of interesting pattern recognition problems including Kanji character recognition,† industrial parts inspection, handwritten character recognition and real-time three-dimensional object classification.

† It was able to recognize 2500 distinct characters. This compares favorably with the average Japanese recognition of 2000-3000 characters.

Neural network models perform comparably to conventional approaches in signal processing. Adaptive filters can be used to recover waveforms from noisy channels and a three-layer network does better than a linear adaptive filter for time series prediction.

For the analysis of many inputs in real-time, the inherent massive parallelism of neural networks can be used. Vision and speech recognition are examples where large amounts of data are preprocessed in real-time but application to more conventional process monitoring has also been achieved.

Many large defense-oriented companies including TRW, Hughes, Honeywell, and Lockheed have active neural net research programs.

DuPont has an active program in the application of neural networks to chemical process control. I am attempting to get more detailed information at this time.

-There is also an active joint European Esprit project for the application of neural networks to industry. The $6 million project has participation from 10 firms in four countries.

5. Application Selection Criteria

Given the above information, it seems desirable to attempt a list of criteria that constitute a filter for the selection of problems and applications where a neural network approach can be attempted. This list is based on current capabilities and will be modified and extended as the field evolves. In looking at your application, the following questions may help to identify problems which might be candidates for a neural network solution:

1. Are you presently using statistical pattern recognition techniques and need to improve their performance?
2. Are there areas where a computer could do the control or analysis but is not doing so because the programming is too complicated or for which a model or algorithm doesn't exist or is not satisfactory?

3. If you have a particular problem, can it be re-cast as a pattern matcher or recognizer that could be trained on historical data which is available?
4. Do you have jobs that require human analysis or judgement that result in a discrete set of pre-defined answers?
5. Do you have problems that you think could use an expert system but don't want to invest the time and programming involved? Many of the problems that need expert systems can be reduced to fundamental problems of pattern classification and association.
6. Do you have a large amount of machine-readable data that you would like to analyze for patterns and associations but haven't gotten around to it yet?

If you answered "yes" to any of the above questions, then you might have a good candidate for a neural network application.

6. A Modest Proposal

Neural networks are not going to produce an R2D2 or C3PO in the near future, but given all of the present activity, they will undoubtedly have an impact on computing practices in the years to come. Many companies presently have or are working on VLSI implementations of massively parallel architectures. It seems that SLAC should have some kind of plan to be involved with this technology and take advantage of its strengths. As the hardware implementations of large networks become available, one should be able to transpose software simulations into hardware, in the same manner as programmable logic devices are handled today.

The initial effort during FY90 should be to select a few applications which can be implemented using well-known, standard models. The problems should be interesting enough that their solution has real benefit but not so interesting that they can't be implemented with conventional hardware in a "reasonable" development and execution time.

Assuming some modest level of success with the selected applications, FY91 should see the expansion into other application areas and the evaluation of the current crop of development systems, perhaps selecting one for standard use at SLAC in the development of neural network applications.

If by FY92 there are at least several successful neural network applications being developed or maintained, the integration of these tools into the standard systems development environment should be started. If appropriate, this would also be the time to start more theoretical work on some "harder" problems which may be amenable to solution with commercial hardware and new models.

REFERENCES

1. *DARPA Neural Network Study*, October 1987 - February 1988, AFCEA International Press, ISBN O-916159-17-5.

2. Rumelhart, McClelland *Parallel Distributed Processing.* Cambridge MA: MIT Press, 1987.

3. Wasserman, Phillip. *Neural Computing, Theory & Practice.* Van Nostrand -Reinhold.

4. Franklin, J. Sutton, R. & Anderson, C. "Application of Connectionist Learning Methods to Manufacturing Process Monitoring," Proceedings of the Third IEEE International Symposium on Intelligent Control, August 1988.

5. Miller, T., Sutton, R., & Werbos, J. *Neural Networks for Control.* MIT Press, Not published yet.

6. J. Hopfield and D. Tank, "Computing with neural circuits," Science, vol. 233, pp. 625-633, August 1986.

7. Widrow, B. & Stearns, S. D., *Adaptive Signal Processing.* Prentice-Hall, NJ, 1985.