CHAPTER 2

INTRODUCTION TO ARTIFICIAL NEURAL NETWORKS

1. INTRODUCTION

A general introduction to the subject of Artificial Neural Networks is given in this chapter and the tenuous relationship of neural networks to the biological neuron structure of the brain is also briefly outlined. The development of artificial neural networks has been marked by periods of considerable optimism and others of disillusionment. A realistic assessment of the potential of artificial neural networks is attempted and some of the unrealistic expectations which has grown around a new and developing subject are dispelled.

The study of artificial neural networks originally grew out of a desire to understand the function of the biological brain. This relationship between the biological neuron and the artificial neuron has been of great importance in past research but at the present time it does not appear to be a fruitful field. A basic description of the biological neural network is included so that the debt the artificial neural network owes to the biological neural network may be appreciated.

Most of problems at now able to be solved by means of artificial neural networks are generally also able to be solved by alterative methods. The question whether to use artificial neural networks to solve a particular problem is a matter of judgement on the part of the designer responsible for the project. The neural network would be a suitable candidate for use if significant advantages in important areas such as cost, speed of operation, reliability, ease of maintenance, ease of initial development, ease of deployment and modification can be shown to exist.

As neural network applications are still in the early stages of development many practical problems are likely to be pioneering applications and it will not be possible to rely on established precedents as a guide. There will therefore be some element of risk in the choice of a neural network should the application fail. The consequences of electing to use neural networks for new applications should neural networks scheme fail to provide the required degree of performance would involve costs such as the loss of time and development costs.

The underlying reason for using a artificial neural network in preference to other likely methods of solution is that there is an expectation that it will be able to provide a rapid solution to a non trivial problem. Depending on the type of problem being considered there are often satisfactory alternative proven methods capable of providing a fast assessment of the

0-8186-7085-1/95 $04.00 © 1995 IEEE

situation. Alternative fast methods of assessment may include such procedures as the Liapunovs Method, the use of a Knowledge Based type of system or perhaps a simplified model of the system would be appropriate.

The use of a Lookup Table Method where all system data has been accumulated would probably not be an acceptable method as by definition the problem is of a non trivial nature. The artificial neural network also has a very great advantage over the look up table approach in that it has the ability to 'generalise'. Generalisation is the ability which a trained neural network has to produce an appropriate answer when presented with previously unseen data. There are restrictions on the scope of the unseen data in that it must not be too different from the data on which the neural network has been trained. Look up tables do not have this ability.

The Artificial Intelligence field as commonly defined is briefly outlined in a later section of this chapter and the relationship of the artificial neural network technique to the Knowledge Based System is considered.

2. LIMITATIONS

To avoid any future disillusionment it is desirable to state at an early stage just what are reasonable expectations for Artificial Neural Networks (ANN). They may be regarded as performing as very good multi dimensional interpolators. In this context their performance is limited by the boundaries of the information submitted to them during the training phase of their development. That is, the bounds of the information provided during the training or learning phase must extend to cover the entire region of anticipated future interest. What is meant by the training phase will be fully discussed in later chapters. On this basis it would be unrealistic to expect a trained ANN to provide intelligent responses to questions which lie beyond the extent of the outer bounds imposed by the extreme limits of the training data.

The trained ANN has the obliging and sometimes irritating property of always providing an answer to any correctly phrased query. Unfortunately the answer provided may be nonsensical if the query is inappropriate. Only the subsequent processing will reveal the extent of the error and the user has a responsibility to refrain from asking questions beyond the boundaries of the ANNs range of training.

Artificial Neural Networks are not universal panaceas to all problems as they are really just an alternative mathematical device for rapidly processing data. It could be argued that animal and human intelligence is 'only' a huge extension of this process in that biological systems learn and then interpolate and extrapolate using the information which has been slow when compared to the speed of an electrical signal, in the order of 300 feet per second in the biological system compared to 186,000 miles per second in the electronic system. Despite this low signal propagation speed the brain is able to perform splendid feats of computation in everyday tasks. The reason for this enigmatic feat is considered to be the degree of parallelism which exist within the brain. It is an interesting sidelight on current technology that due to the serial nature of the normal von Neumann type computer and programs available it is necessary to treat the intrinsically parallel artificial neural network as a serial process. This computing contradiction will be resolved when facilities for parallel computation become commonly available.

The method of operation of the brain has long been a source of interest to scientists but owing to the obvious difficulties in studying the living brain little progress was made in the understanding of the manner in which a brain functions. Laboratory animals have been used for many years in experiments on non-human brains and cases of brain trauma caused by injury have provided opportunities to gain some insight into human brain function. Some authorities consider that the complete understanding of the brain will take a very long time to achieve if ever.

NEURON ARRANGEMENT

The fundamental building block of the brain is the neuron and a simplified representation of a biological neuron is illustrated in Figure 1. The shape of each neuron is distinct as are all the other component parts of each neuron. Each neuron is therefore unique as no two neurons are ever identical. It is estimated that the human brain contains somewhere in the order of 10^{11} neurons. The neuron is comprised of three distinct sections and each of these sections is now briefly considered.

CELL BODY

The cell body is approximately spherical or roughly pyramidal in shape and it contains the cell nucleus and the mechanisms necessary for the survival of the cell. While every cell of a particular kind is an individual and has its own particular arrangement it is still generally similar in shape and arrangement to other cells of the same type.

DENDRITES

The dendrites form the physical means by which incoming information is received by the cell. The dendrites are arranged in a tree like structure which has many branching processes.

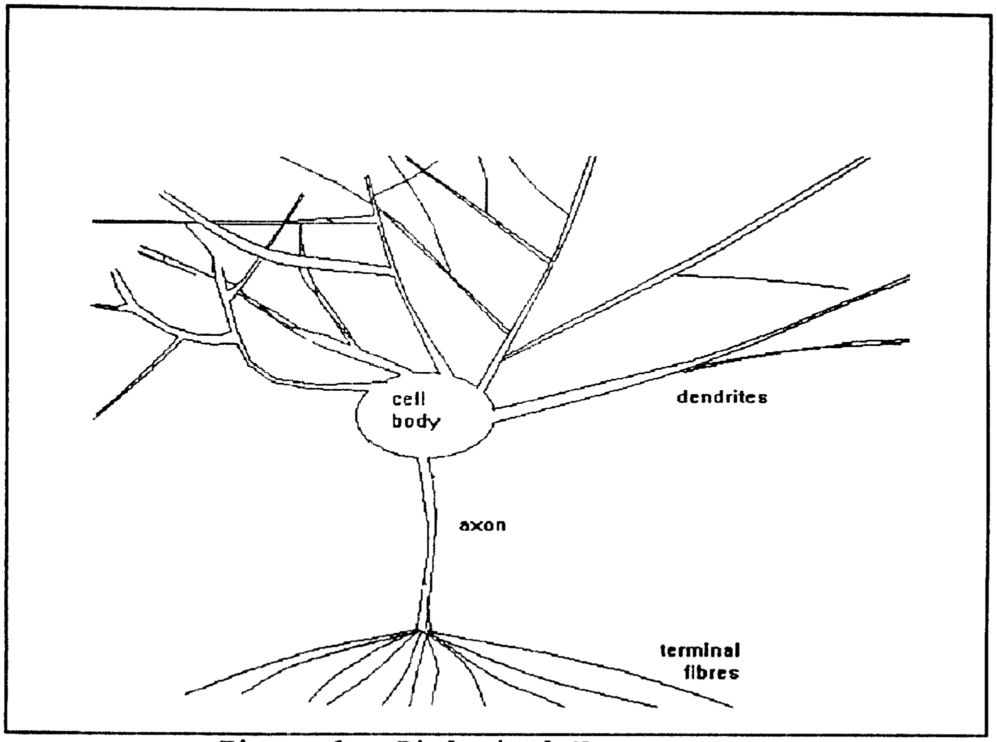


Figure 1. Biological Neuron

previously learnt. Neural Networks appear to have the property of interpolation but not that of extrapolation.

The name Artificial Neural Networks given to the study of these mathematical processes is in a sense unfortunate in that it creates a false impression which leads to the formation of unwarranted expectations. Despite some efforts to change to a less spectacular name such as Connectionist Systems it seems that the title Artificial Neural Networks is destined to remain. At this time the performance of the best ANN is trivial when compared with even the simplest biological system.

3. BIOLOGICAL INFLUENCE

The study of Artificial Neural Networks initially grew out of attempts to understand the working of animal physiology. The basic drive which led to the development of ANN was a desire to understand the behaviour of the brain. The first attempts to model the behaviour of individual neurons in the 1940's were hardware based as these early studies predated the ready availability of the digital computer.

The speed of signals in the biological network are intrinsically very

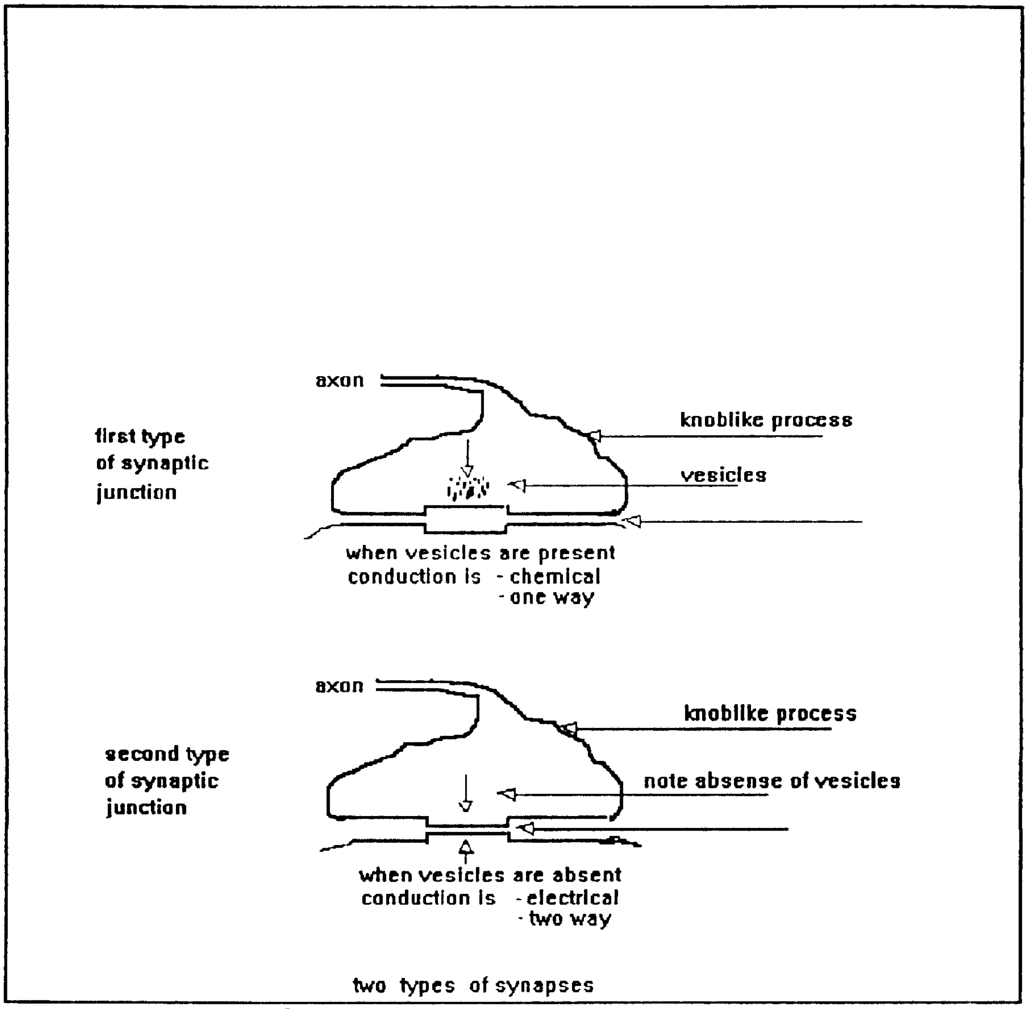


FIGURE 2. SYNAPTIC JUNCTION

AXON

The axon is the output channel of information from the neuron and extends away from the body of the cell. The axon divides into a series of processes which in turn make connection with other neurons in the interconnected system. The axon is longer than the dendrite processes and in some instances may be a metre or so long.

The components of a typical model of a single typical neuron have now been approximately defined and it is now necessary to see how a number of such neurons can be connected together to form a cluster of cells.

SYNAPSES

Connection between the axons and the dendrites of individual neurons is the most frequent form of interconnection between neurons in the brain. While axon-dendrite connections are the most common form of connection, other forms such as dendrite-dendrite connections and axon-axon connections have also been observed.

The connection of these elements of different neurons occurs at a specialised structure called a 'Synapse'. A synapse is the point where information is transmitted from one neuron to another neuron by means of a chemical action or the transfer of an electrical signal. The tip of the axon has been described as being 'Knoblike' and this structure contacts the receptor section of the other neuron. The two elements of the synapse are separated by a region called the 'Synaptic Cleft' The synaptic cleft in some types of synapses is a very thin band of about 200 nanometres thickness. The transfer of information between the cells occurs at this point. A number of different types of synapses are known and two commonly occurring types of synapses are shown in Figure 2.

CONNECTION OF NEURONS INTO GROUPS

The neurons in the brain are connected together in an exceedingly complex manner and it is estimated that there may be in excess of 10^{14} connections between neurons. This figure assumes there could be 10^{3} connections for every neuron if the connections were evenly distributed. All estimates of numbers and quantities concerning neurons are considered to be indicative and are only estimates.

Supposing complete interconnection between all biological neurons were to be attempted then there would need to be something like 10^{22} separate connections to each neuron. Such a level of connection would be clearly impossible to attain even for the brain. It is considered by some that each axon can form up to 1000 connections. If this figure were to be accepted then an upper limit of possible connections would be 10^{14}. It can be seen that although the brain is heavily interconnected the connection matrix when compared to an electrical transmission system would probably still be considered to form a relatively sparse matrix.

The comparison between the biological neuron and the mathematical or electronic neuron is possible by comparing Figure 1 with Figure 3. Such a comparison is very interesting but it is probably not directly useful the application of mathematical neural network to practical problems. For this reason it is not intended to pursue the similarity of the two kinds of systems any further in this book. The remainder of this book will therefore be restricted to a consideration of the mathematical neuron.

4. HISTORICAL SUMMARY

Artificial Neural Networks now have a relatively long history and a correspondingly large amount of literature exists on their properties and development. It would be unrealistic to attempt to provide an extensive coverage of the literature. The study is accordingly restricted to some of the most significant historical markers in the development of ANN's and these are given below. The references and events outlined in this section cover only what are believed to be major developments and turning points in ANN research.

McCULLOCK AND PITTS

The first significant paper on artificial neural networks is generally considered to be that of McCullock and Pitts in 1943. This paper outlined some concepts concerning how biological neurons could be expected to operate. The neuron models proposed were modelled by simple arrangements of hardware which attempted to mimic the performance of the single neural cell.

HEBBS

The book 'The Organisation of Behaviour' written by Hebb in 1949 formed the basis of 'Hebbian Learning' which forms an important part of ANN theory today. The basic concept under lying 'Hebbian Learning' is the principle that every time a neural circuit is used, the pathway is strengthened.

AVAILABILITY OF DIGITAL COMPUTERS

About this time of neural network development the digital computer became more widely available and its availability proved to be of great practical value in the further investigation of ANN performance.

JOHN VON NEUMANN

In 1958 Neumann wrote a book 'The Computer and the Brain' in which he proposes modelling the brain performance by items of hardware available at that time.

FRANK ROSENBLATT

Rosenblatt constructed neuron models in hardware during 1957. These models ultimately resulted in the concept of the Perceptron. This was an important development and the underlying concept is still in wide use today.

BERNARD WIDROW AND MARCIAN HOFF

The researchers Bernard Widrow and Marcian Hoff were responsible for the

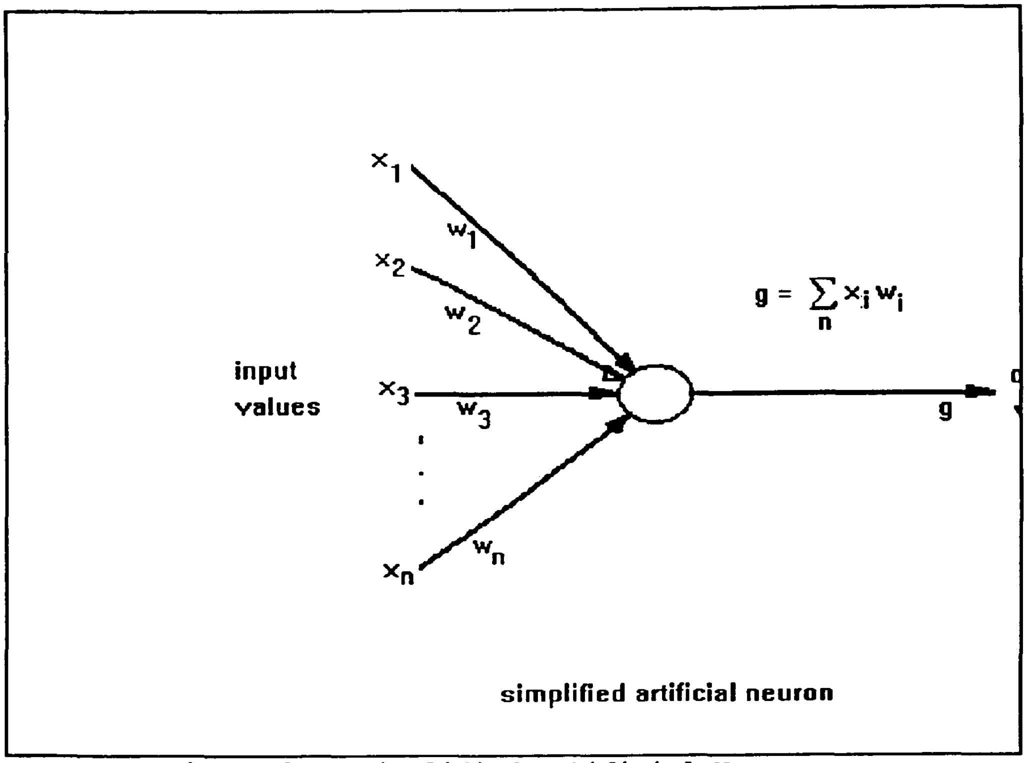


Figure 3. Simplified Artificial Neuron

development of first the ADALINE and then the MADALINE networks. The name ADALINE comes from 'ADAptive LINEar complier' and the name MADALINE comes from 'Multiple ADALINE' respectively. Much was made of the potential scope of the technique but this was succeeded by a period of disillusionment.

MARVIN MINSKY AND SEYMOUR PAPPERT

In 1969 Marvin Minsky and Seymour Pappert published an influential book 'Perceptrons' which showed that the perceptron as developed by Rosenblatt had serious limitations. This very thorough work was well researched and showed that the Perceptron in the form it had at the time suffered from severe limitations. The essence of the book Perceptrons was the assumption that the inability of the perceptron to be able to handle the 'exclusive or' function was a common feature shared by all neural networks. As a result of this assumption interest in neural networks was greatly reduced. The overall effect of the book was to reduce the amount of research work on neural networks for the next 10 years. The book served to dampen the unrealistically high expectations previously held for ANN's. Despite the reduction in ANN research, a number of people still persisted in ANN research work.

JOHN HOPFIELD

After 10 years in the doldrums John Hopfield produced a paper in 1982 which showed that the ANN had potential for successful operation and showed how this could be developed. This paper was timely as it marked a second beginning for the ANN. While Hopfield is the name frequently associated with the resurgence in interest in ANN it probably represented the culmination of many peoples work in the field.

From this time onwards the field of neural computing began to expand and now there is world wide enthusiasm as well as a growing number of important practical applications.

5. ARTIFICIAL INTELLIGENCE

For completeness the associated areas of artificial intelligence will be outlined. The subject of artificial intelligence is generally considered to include the areas of Knowledge Based Engineering, Robotics, and Artificial Neural Networks. However the term is not strictly defined and very often other related areas such as Genetic Algorithms are included under the heading of artificial intelligence.

Knowledge Based Engineering is based on the assembly of a large number of rules relating to a particular subject to form a data base. A complete set of such rules is then considered to contain all the information relevant for the system under consideration and is then capable of adequately dealing with any query regarding the system being studied.

The set of rules is obtained as the result of interrogation of one or a number of people considered to be expert in the area under consideration. For a problem of any significant size, that is a non-trivial problem, the number of rules required avalanches and soon becomes very large indeed. The enormous size that a data base may easily reach causes interrogation to become a lengthy or possibly even an infinitely long process. A rule to rule search of such a data base by this or any other non intelligent method is therefore impracticable for any but the smallest system.

The interrogation of a large set of rules in the most advantageous manner possible presents a difficult problem. Most systematic methods of data base search seem to assume a tree like structure and this must be searched systematically until either an acceptable solution is found or until it can be eliminated from further study. A necessary property for a systematic data base search routine must include some device to eliminate any branches which have already been consulted and previously rejected. Programs to extract information from the data base may be written in any computer language but they are normally written in a specialised language such as LISP or PROLOG due to the ease with which these languages can process symbolic information.

6. PATTERN RECOGNITION

Pattern recognition may be considered to be the attempt to identify parameter relationships existing within sets of assembled data. The pattern recognition process consists essentially of two steps. The first step being the collection of data and the second step the extraction of the relationships existing between the parameters contained within the assembled data. This two step scheme is a simplified version of the actual pattern recognition process and in practice the extraction would generally be more complicated. Particular attention would need to be given to the assembly and preprocessing of the data before the feature extraction process could be attempted. It has been said that if either the processes were perfect then the other procedure would become trivial.

The feature extraction process has traditionally been based on some form of statistical analysis to obtain the best fit of the parameters. From one point of view the ANN may be regarded as a different method of feature extraction. These two steps of data preparation and feature extraction are present when the artificial neural network procedure is adopted.

Table 1 indicates the differences existing between the approaches used when employing artificial neural networks, expert systems and conventional computing techniques. It can be seen from the table that significant differences exist between the three approaches.

7. WHY USE NEURAL NETWORKS ?

The resurgence of interest in artificial neural networks over the last few years is due to a number of factors. One of the continuing and significant driving forces in neural network research has been the desire of many diverse groups which include neuro physicians, engineers, psychiatrists, psychologists and biologists to gain an understanding of the workings and behaviour of the brain.

Some of the valuable features of artificial neural networks which distinguish this method of computation from other algorithm based methods of computation are now considered.

TABLE 1. A COMPARISON BETWEEN NEURAL NETWORKS, EXPERT SYSTEMS AND CONVENTIONAL PROGRAMMING

Parameters	Artificial Neural Networks	Expert Systems	Conventional Programming Techniques
Process	Learning	Inference	Algorithm
Input Data	Pattern	Knowledge	Numerical
Algorithm	Statistical	Heuristic	Programming
Computation	Numerical	Symbolic or Logic	Arithmetic or Logic
Data Processing	Parallel	Serial	Serial
Output Results	Inductive	Deductive	Computed

THE GENERALISATION CAPABILITY OF NEURAL NETWORKS

One of the most important if not the most important attribute of neural networks is the ability to generalise. By this is meant the ability of a neural network to succesfully interpret data which it has not previously encountered and to provide a sensible result. It is this property which sets the neural network in a different category to a system such as a look up table where it is necessary to store all of the information likely to be required for reference on future occassions.

Naturally there are limits to the generalising ability of any network and it is essential for the network to have been trained on information which is closely related to that on which the network is expected to generalise. In the sense of generalisation it is helpful to consider the network as an interpolater within a multidimensional space. It is very different to the normal process of interpolation which is carried out by means of an algorithm where as here the information has been learnt by the trained network.

The extrapolation ability of a network beyond the boundaries of the training information provided is non existant. The network has no way of having this information.

PARALLELISM IN APPROACH

Convential programable computers which operate in a serial manner as proposed by von Neumann appear to have reached a performance plateau based on the limits set by the switching speed of the present component elements. One way of overcoming the barrier imposed by component speed limitations is to develop a different form of computer which is able to operate in a non serial manner. Such a computer would enable operations to be performed in parallel rather than having to wait for one operation to be completed before another operation can commence. Computers based on these principles are being developed but at the present time neither the computers or suitable programs are widely available.

An alternative method of increasing the speed of calculations is to utilise a neural network which works in a simulated parallel manner and is not limited by the serial requirements of the normal program. There is a slight paradox at the present time in that the parallel functioning neural network is implemented by means of the normal serial computer. The neural network which is an intrinsically parallel functioning device is forced by the limitations of current computers into what is a serial functioning mode of operation. This situation will in all probability be rectified when parallel operating computers and appropriate programs become more generally available.

DISTRIBUTED MEMORY

One of the advantages claimed for a neural network is the fact that the memory is distributed over a large number of components within the network. This has been likened to the performance of the biological brain where a degree of damage may be sustained without an immediate reduction in ability. The loss of neurons within the brain during the natural process of aging does not result in an immediately obvious loss in the organisms performance.

The robust nature of the neural network is obtained as the result of many parallel paths being available for use. In the case of the artificial neural network in the interests of training efficiency there is an imperitive to reduce the number of neurons existing in any network. The reduction in neuron numbers while it improves the training performance results in a reduction in the robustness of the network to damage.[]

NEURAL NETWORKS EXHIBIT INTELLIGENT BEHAVIOUR

The trained artificial neural networks are sometimes said to exhibit intelligent behavior. Statements of this type raise the question as to what is meant by intelligent behaviour in these circumstances? In the case of a living organisium it could perhaps be said to be behaviour which improves the organisms chances of survival by finding food, finding a mate or possibly eluding the evil designs of a predator.

None of these conditions fit the case of a neural network and some other measure must be adopted. Possibly [Turings] concept of an inquirer being unable to tell in a 'blind' type interogation if the unseen respondant is a person or a machine may be a suitable criterion. At the present time practical neural networks are very specialised and any interogation of this kind would need to be strictly limited to the networks restricted area of expertise. It could be said that the limitation of the inquiry is only a matter of degree because even an interogation of an individual must also be limited to those matters within the individuals range of experience.

LEARNING NOT PROGRAMMING

The neural network is not programmed during training but actually learns the information being presented to it. At the present time it does not appear to be possible to readilly interpret the information stored within the trained neural network system. Understanding of this information may become possible in the future when a more mature understanding of the operation of neural network theory is developed.

Conventional computers do not have the ability to learn and consequently do not have the ability to generalise to accomodate previously unseen data.

GENERAL ATTRIBUTES OF NEURAL NETWORKS

In the real life situation there are many areas where expertise is required in order to accomplish the desired result. Many of these areas of interest involve assessment of almost continuous operations such as, the supervision of a power system operating condition, or the testing for explosives of baggage belonging to embarking air travellers. Work of this kind when done by human beings is highly repetitive and consequently boring factors which lead to operator inattention with an inadvertent lowering of surveillance standards. In addition the cost of keeping a sufficiently large crew of trained staff on hand to continuously monitor processes such as this is extremely expensive. The neural network has the potential in some situations to provide an alternative method which is tireless, continuous, reliable and inexpensive replacement of personal for routine work of this kind.

The artificial neural network may be relied on to undertake suitable designated tasks in a systematic manner at a speed which could not be attained or maintained by human operator. Such advantages naturally involve development costs and very likely some continuous maintenance costs.

8. BRIEF REVIEW OF NEURAL NETWORK BASIC COMPONENTS

The range of types of neural networks which have been developed is large and it is desirable to attempt to establish some kind of relationship between the various kinds of networks. There are a number of ways in which neural networks may be categorised based on characteristics such as,

- -The method of training adopted, directed or non-directed
- -Whether after training feedback or non feedback operation is involved,
- -The type of training algorithm employed,

In order to consider the operation of artificial neural networks it is first necessary to introduce some of the terms used. This will be done in the following section.

NEURONS

The neuron forms the node at which connections with other neurons in the network occur. Like the biological network, the neuron in the artificial network is also central to network operation as much of the activity on the system occurs at the neuron. Althougth the infinitely more successful biological neural network neurons are not arranged in any geometric pattern those in the electronic network are generally arranged in one or more layers which contain neurons performing a similar function. Depending on the type of neural network being considered, connections may or may not exist between neurons within the layer in which they are located. For example in the Back Propagation Network there are no connections between the neurons in the same layer but in the case of the Hopfield Network every neuron is connected to all neurons in the layer.

WEIGHTS

In the trained artificial neural network the intelligence of the network is stored in the values of the connections existing between the neurons. In artificial neural network terminology the values of the connections between the neurons are generally referred to as weights.

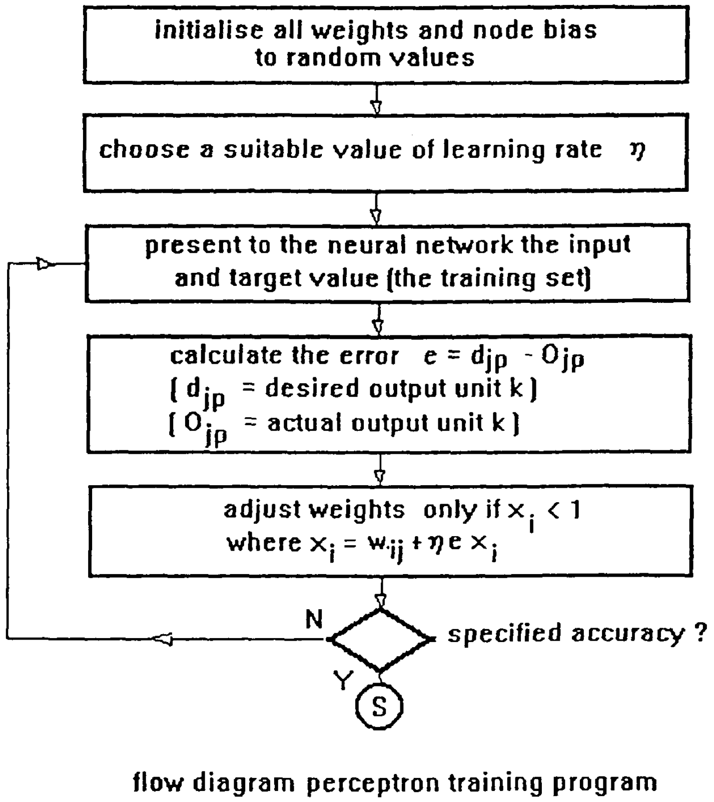


FIGURE 4. FLOW DIAGRAM PERCEPTRON TRAINING

hidden layers also take part in producing output when training is complete and the network is being interrogated.

The number of hidden layers provided is problem dependant. There may be advantages in providing several hidden layers but additional layers may mean a marked increase in the training time taken.

9. TRAINING (LEARNING)

In contrast to expert systems which incorporate a knowledge base, neural networks do not have such a collection of information. They need to be trained for a given problem or situation so that the weights will then contain the required information. An example of the iterative proceedure necessary during training is given in the percepron training flow diagram Figure 4. One of the ways of classifying training proceedures into two categories is whether directed training or non directed training is employed and these methods are now considered.

SUPERVISED TRAINING or DIRECTED TRAINING

When employing directed training it is necessary to include among the set of data presented to the neural network the result or answer corresponding to each particular set of data. The data set is repeatedly presented to the neural network until the network output corresponds closely enough to the result in the data set. Should the difference between the network output and the 'Target Value' exceed the permitted tolerance the network training process is repeated. The algorithm being used for the network training causes further adjustment to occur and the process is repeated until the tolerance is acceptable.

In supervised training training data set contains a collection of information representing the input pattern vector together with the target value or desired output. This set of training data is presented to the network repeatedly until the difference between the target output and the actual output of the network reaches a certain predetermined value. The perceptron training flow diagram given in Figure 4 uses directed or supervised training and the proceedure for obtaining the difference value may be seen in the figure.

After each evaluation of the training algorithm the difference value between the actual output and the target output contained in the training set is compared with the permitted error value. If the difference is equal to or less than the error value when the process is stopped then the network is considered to be trained. Should the difference be greater than the permitted error value the set of training data is again presented to the training algorithm and the connection weights of the neurons changed until the output satisfies the criterion.

UNSUPERVISED TRAINING or NON DIRECTED TRAINING

In the case of non directed training the target value, or answer, is not provided and the information in the training data set is continuously presented until some convergence criteria is satisfied.

In unsupervised training the output pattern for a given input pattern is not provided. The neural network constructs internal models that capture regularities in input pattern.

For any problem it is necessary to provide the neural network with training data which covers the extent of the problem. This data will need to include sufficient information so that the problem is unambiguous. The training set is repeatedly applied to the neural network until some specified training criteria is met.

SELECTION OF TRAINING DATA

The selection of training data is a crucial and difficult task with conflicting requirements which must be satisfied in the best way possible.

The first requirement is that there is sufficient training data presented to the network so that an unambiguous output is able to be obtained. This requires the careful selection of parameters for the particular problem. In some cases the availability of sufficient suitable data may be limited and this can pose a difficult task in itself.

The elimination of any data which does not contribute to the performance of the trained network is also required. This may be done automatically by eliminating any parameters which are not participating in the training process. Caution must also be exercised in this respect as some data may seem to be contributing little to the solution but when eliminated the network performance is significantly degraded.

TERMINATION OF TRAINING

The decision when to stop neural network training is a difficult one. If the problem is a relatively simple one it is easy to say when the answers provided during training and testing fall within certain predetermined error limits.

However significant practical problems are rarely as clear cut as this. If an RMS error criterion is used this does not eliminate relatively large errors occurring with individual training sets even when acceptably small RMS errors are obtained. In some cases this may not matter but in many instances the occasional grossly incorrect answer may be unacceptable. Also if nice answers are provided always the problem may be a simple one and perhaps simpler more conventional alternative methods are available in preference to a neural network.

10. FEED FORWARD AND FEEDBACK NETWORKS

Feed Forward Networks are networks which after completion of the training process provide answers to queries as the result of a single forward pass of the data through the network. However during the network training phase a feed forward network may be traversed many times before the training criteria are satisfied and the network is considered properly trained. Once the training process is complete interrogation of the feed forward trained network requires only a single pass.

The training process for feedback networks is similar to that of feed forward networks but differs during the interrogation process. Even when training is complete a feedback network will require a series of passes during the interrogation in order to provide an answer to a query.

An example of a feed forward neural network is the Back Propagation network while the Hopfield network provides an example of a feedback network.

11. HEBBIAN LEARNING

The concept of this form of learning was stated by the psychologist Donald Hebb in his book [Hebb 1949]. Essentially the proposition is that every time a neural pathway is used it is reinforced. This approach is known as Hebbs Learning Rule or Hebbian Learning and is still central to many of todays neural network learning algorithms.

The association of the learning process with the cell is significant as it appears to roughly approximate the actual biological learning process. There is evidence that the Hebbian learning rule actually does apply to biological neural networks. Although the Hebbian learning rule appears to apply to the actual biological learning process it is thought to be much more complex than that suggested by this rule.

It is one of the simplest ways of learning and consists of purely feed forward unsupervised learning.

12. THRESHOLD FUNCTIONS

A threshold function provides a means of further processing the output of a neuron after the initial processing has taken place. It enables signals of greatly differing amplitudes to be satisfactorily dealt with in the neuron. Many different functions such as the sigmoid and the tanh have been utilised for this purpose. Both of these functions provide the following desirable characteristics,

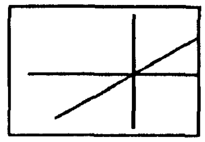
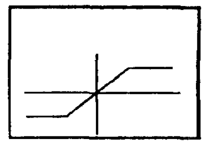
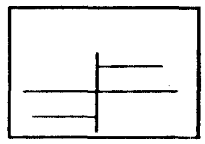
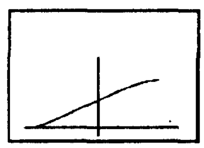
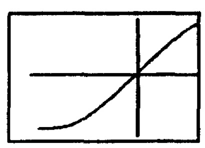
(1) small signals are dealt with in a linear manner

(2) large signals are limited to a maximum value.

The sigmoid function is favoured as it has a simple mathematical form which in certain circumstances enables the analysis to be simplified.

Threshold functions are also called activation functions, signal functions, or squashing functions are used to map the input pattern of a neuron to the specified output range. The most commonly used threshold functions are listed in Table 2.

TABLE 2. MOST COMMONLY USED THRESHOLD FUNCTIONS

LINEAR FUNCTION (Ramp Function) This function - has a linear zone - is easily implemented	$f(x) = \alpha x$ where α is a constant 
NON LINEAR (RAMP) FUNCTION This function is used to represent simplified non linear operation	$f(x) =$ 1 if $x \geq 1$ x if x (1 -1 if $x \leq 1$ 
STEP FUNCTION (HARD LIMITING) This function is - fast and easy to implement - has no linear range - cannot smoothly imitate functions	$f(x) =$ 1 if x) 0 -1 otherwise 
SIGMOID FUNCTION (S SHAPED FUNCTION) This function is - continuously differentiable - can make fuzzy decisions - not easy to implement - the output is limited to positive values	$f(x) = 1 / (1+e^{-1})$ 
HYPERBOLIC TANGENT FUNCTION	$f(x) = \tanh(x)$ 

13. ARTIFICIAL NEURON

Figure 5 illustrates a single artificial neuron which is a building block of all the neural networks. The quantities $x_1, x_2, \ldots x_n$ is the

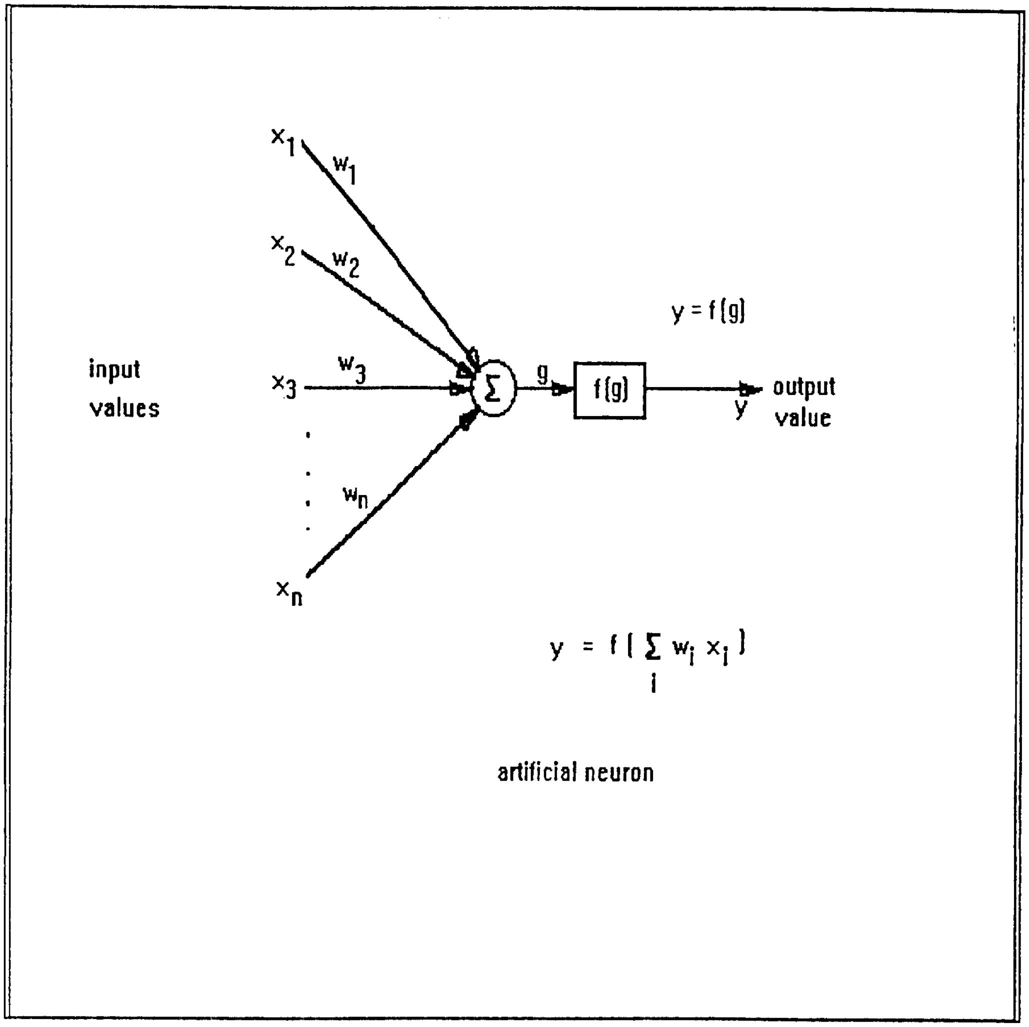


FIGURE 5. ARTIFICIAL NEURON

set of vector inputs to the neuron. These inputs correspond to the stimulus level of neurons within the brain. This input vector is applied to the neuron through a set of associated weights $w_1, w_2 \ldots w_n$ and these correspond to the strength of the connections within the biological neuron synaptic. The total output of this section of the neuron is given by the expression:

$$g = x_1 w_1 + x_2 w_2 + \ldots + x_n w_n \quad (1)$$

$$= \mathbf{X}\,\mathbf{W}$$

This output is further processed by a threshold function f() to produce the neuron's output as illustrated in Figure 5. There are various kinds of threshold functions as described previously in Table 2.

A simple example of a threshold function is:

$$f(y) = 1 \text{ if } y > 0 \qquad (2)$$
$$= 0 \text{ otherwise.}$$

It is interesting to see the analogy between electronic circuits and brain provided by Hopfield and shown in Table 3. Some comparisons between a number of important features are shown in this table. For example the degree of connectivity in the biological network is very much larger than that in a typical electronic network by a factor of 1000 to 3.

TABLE 3. QUALITATIVE HOPFIELD'S COMPARISON BETWEEN THE HARDWARE OF PRESENT ELECTRONIC CIRCUITS AND THE WETWARE OF NEUROBIOLOGY

Features	Electronic Circuits	Brain
Connectivity	Approximately 3	3000
Connection Matrix	Mainly feed-forward	Feedback
Mode	Digital / Analog	Analog
Timing	Clock	Dynamic System
Delays	Suppressed for speed	Used for Processing
Circuit	Fixed	Variable

In the electronic circuit the operation is predominantly feed forward while in the biological system there is a high degree of feedback as part of the process of calculation if so mundane an expression may be applied to brain activity.

In the biological system delays are used as part of the process while in the electronic system they are minimised in pursuit of higher speeds. The biological system is completely analog and automonous.

14. PERCEPTRON

Frank Rosenblatt in 1958 introduced a neuron like artificial learning device called perceptron.

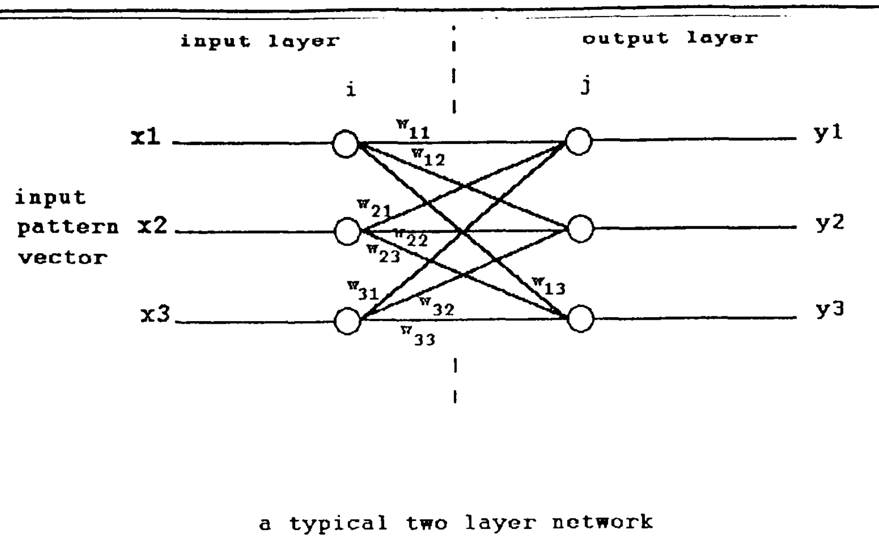


FIGURE 6. TYPICAL TWO LAYER NEURAL NETWORK

Figure 6 illustrates a typical two layer network. The input layer of neurons have three inputs x_1, x_2 and x_3. Each of these inputs can assume either the value of 0 or 1. The input layer feeds its output to the output layer. The neurons in the output layer do all the necessary processing and deliver the output. All the neurons in the input layer are connected to all the neurons in the output layer. This network is also called a fully connected neural network.

The output of the network is expressed as:

$$y_j = \Sigma_i \; w_{ij} \, x_i \tag{3}$$

where

w_{ij} = connection weights from neuron in the input layer i to the neuron in the output layer j.

It can be shown that these two layer networks are limited in applications. It is possible to include one or more hidden layers between the input layer and the output layer. It will make the network general but will not necessarily produce the best results.

15. ADALINE

In 1959 Widrow invented a basic building block called Adaline for implementing the neural networks. Figure 7 illustrates the simplified representation of Adaline. It consists of an adaptive linear combiner cascaded with a hard limiting quantizer. During the training the network is presented with the input patterns and the target output response. The difference between the target output and the actual output also called an error is fed back to the input of the Adaline to adjust the weights. This training procedure is repeated till the difference between the output response and the target response reaches to a predetermined value. An adaptation rule such as Least Mean Square (LMS) algorithm is used for adjusting the weights of the Adaline. Once the network is trained it presents the correct conclusion to the given input patterns. Adaline could also generalise in the sense that it gave close conclusion to the input patterns that were not included in the training set.

16. LINEAR SEPARABILITY

It can be shown that a two layer artificial neural network structure is not capable of realising all the functions pertaining to a given input pattern. For n binary input pattern there are 2^n possible input patterns. For 2^n input patterns (each pattern can have 0 or 1) there are 2^{2n} possible logic functions connecting n inputs to single binary output. It can be shown that a two layer network is capable of realizing only a small subset of these functions known as linearly separable logic functions or threshold functions. Let us consider the case of an Exclusive - OR function which is not linearly separable. Figure 8 illustrates a truth table pertaining to a two variable function. In this example x_1 and x_2 are the inputs. There are four possible combinations of these inputs. It can be shown that no single straight line exists which can separate the two output responses (0 and 1). These non-linearly separable functions can be implemented by incorporating a layer of neurons (also called hidden layer) between the input layer and the output layer of the network.

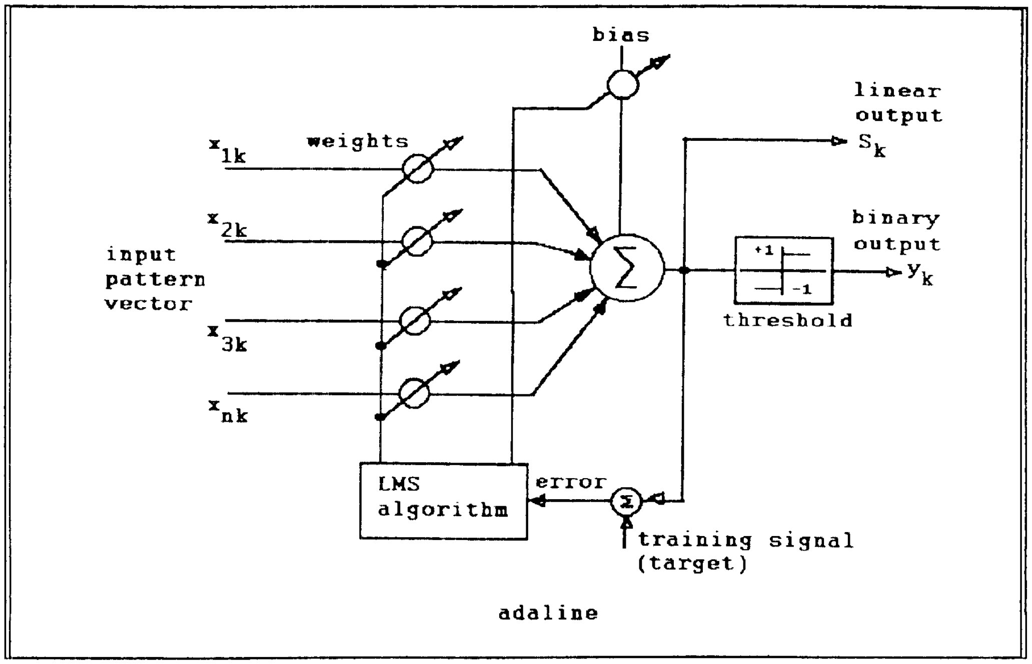
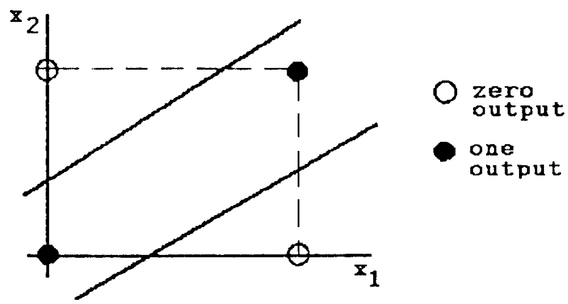


FIGURE 7. ADALINE NEURAL NETWORK

inputs		output
x_1	x_2	$x_1 \oplus x_2$
0	0	0
0	1	1
1	0	1
1	1	0



a exclusive or function

FIGURE 8. THE EXCLUSIVE OR FUNCTION

17. MADALINE

A Madaline (multiple adaptive linear neurons) was also invented by Widrow. Figure 9 illustrates a two-adaline form of Madaline. This structure can implement the exclusive OR function.

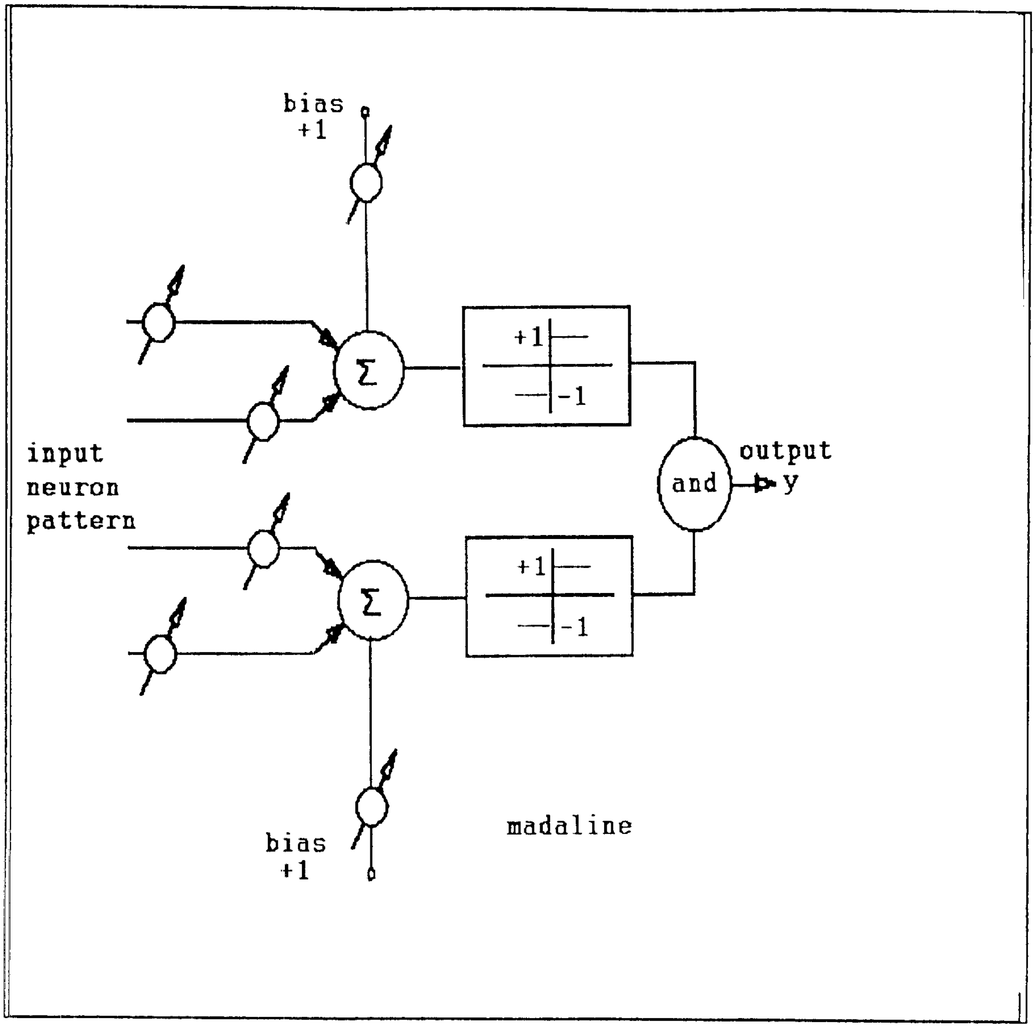


FIGURE 9. MADALINE NEURAL NETWORK

18. CONCLUSIONS

Neural networks operate in parallel and are modelled after the human brain. These networks are increasingly successfully used in various applications including areas as diverse as,

- image processing
- optimisation
- fault diagnosis
- expert systems.
- speech recognition
- radar target recognition
- power system operation

Neural networks are potentially faster in operation than serial computational methods due to the high degree of parallelism and are fault tolerant as information is stored in a distributed form. It is necessary to train neural networks with information pertaining to the given problem. They are able to generalise in the sense that they are able to arrive at a near conclusion to the input patterns that were not taught before. That is subject to the proviso that the question put to the artificial network has been covered in the training set data and is not too different to data encountered during the training process.

There are however a number of problems related to the practical of these networks. For example,

-Some training algorithms fail to converge for a given problem.

-No definitive guidelines available for a network topology required to perform a particular task.

-For real world problems the Networks become very large.

-The training time required may become large.

-No assurance that the answer obtained from the network is correct.

That is wrong answers are occasionally provided by the correctly trained network and due allowance needs to be taken of this fact.

A cursory overview of the biological neural network has been given to illustrate the relationship of the artificial neural network to brain construction and also to emphasise the triviality of the artificial system when contrasted with the biological system.

The most valuable property of an ANN is its ability to learn and to generalise. These attributes mark the neural network out from other methods such as lookup tables. The learning process of the artificial neural network appears to be just a sophisticated multi dimensional interpolation method but then perhaps this may be 'all' normal biological learning.

Anyone seeking an easy solution to a practical problem may be disappointed in their early attempts to make use an artificial neural network. It needs to be remembered that the artificial neural network is only a special mathematical technique and like all such techniques has its limitations. Artificial neural networks appear to have their place and have proved to be very successful in certain selected applications.

A normal characteristic of the neural network approach is the uniqueness of the method adopted for a particular problem. Each practical problem must be carefully considered and the required outputs from the neural network carefully chosen. Every variable included in the problem must be demonstrated to contribute to the improved performance of the neural network.

It is easy to apply too rigid requirements on the output demanded of the neural network. Any undue sophistication in output specification may make the difference between the neural network providing an adequate output for the particular problem or being completely unsatisfactory.

A number of neural networks are used and reported in the literature. Some common examples of different types of networks are,

Perceptron Network

Adaline and Madaline Networks

Back Propagation Network

Radial Basis Function Network

Hopfield Network

Kohonen's Feature Map Network

Adaptive Resonance Theory (ART)

Counter-Propagation Network.

Adaline and Madaline have been discussed earlier in this chapter. The multi-layer perceptron network is an extension of a single layer perceptron network. It has one or more layers usually called the hidden layers between the input layer and the output layer. The Hopfield is a nearest match finding network. This network alters the input pattern through successive iterations until a learned pattern evolves at the output and output no longer changes on successive iterations. The Kohonen self organising features maps organise connections between input neurons similar to that considered to be in the brain. Kohonen proposed that the topological shape of the neuron activation map is similar to the shape of the input pattern. The neural networks based on adaptive resonance theory (ART) are capable of self-organising the stable recognition codes in real-time in response to arbitrary sequences of input patterns. Counter-propagation networks work like look-up table in parallel for finding closest example and its equivalent mapping.

19. REFERENCES AND FURTHER READING

[1] Aleksander, I. and Morton, H., An introduction to neural computing, Chapman and Hall (1990).

[2] Beale, R. and Jackson, T., Neural Computing, Adam Hilger (1990).

[3] Castillo, E. and Alvarez, E., Expert systems : uncertainty and learning, Elsevier applied science (1991).

[4] Dayhoff, J.E., Neural network architectures, Van Nostrand Reinhold (1990).

[5] Diederich, J., (Editor), Artificial neural networks : concept learning, IEEE Computer Society press (1990).

[6] Gelenbe, E., (Editor), Neural networks : advances and applications, North-Holland, (1991).

[7] Hetcht-Nielsen, R., Neurocomputing, Addison-Wesley (1990).

[8] Kosko, B., Neural networks and fuzzy logic, Prentice-Hall (1992).

[9] Lisboa, P.G.J., (Editor), Neural networks, Chapman and Hall (1992).

[10] Simpson, P.K., Artificial neural systems, Pergamon press (1990).

[11] Wasserman, P.D., Neural computing : Theory and practice, Van Nostrand Reinhold (1989).

[12] Winston, P.H., Artificial intelligence, Addison-Wesley Publishing Company, (1992).

[13] Vemuri, V., (Editor), Artificial neural networks : theoretical concepts, IEEE Computer Society press (1988).

[14] Zornetzer, S.F., et. al. (Editors), An introduction to neural and electronic networks, Academic Press (1990).