Psychiatric Bulletin (1991), **15**, 211–213

Computers in psychiatry

Neural networks and psychiatry

CARRICK McDONALD, Consultant Psychiatrist, Warlingham Park Hospital, Warlingham, Surrey CR3 9YR (correspondence); and SIMON McDONALD, Systems Specialist, Torus Systems Ltd, Beaufort Court, Marsh Wall, London E14 9XL

This paper draws the attention of psychiatrists to the developing field of neural networking in the belief that the models created in this discipline demonstrate several functions central to cerebral performance.

Since the brain was acknowledged to be the site of consciousness, cognition, and emotional drives, models have been created to allow researchers to tackle this organ's intrinsic complexities. We mention only the models of Freud, a psychoeconomic model, and more recently a model suggested by Wechsler relating emotional states to brain loci, before turning, without apology, to the general area known as artificial intelligence, which includes the subject of this survey. Boden (1977) has defined artificial intelligence as "the use of computer programmes and programming techniques to cast light on the principles of intelligence in general and human thought in particular". This definition has been refined from the challenge posed by Turing's famous test which stated that if a human communicated with two computer terminals and could not tell which of them was being directly controlled by another (hidden) human and which was being controlled by pre-programmed instructions, then the programme must be conceded to involve "intelligence". Servan-Schreiber (1986) has surveyed the general applications of artificial intelligence to psychiatry. He described the programme ELIZA which, working from an algorithm produced type responses to patients' questions and complaints, which had the general pattern of a psychotherapist employing accurate empathy! He cited an extension of this whereby the "patient" produced paranoid responses, and he talked of expert systems which simply followed a train of reasoning through a series of yes/no gates to lead to a diagnosis. The success of these 'expert' systems is well documented, and Servan-Schreiber speculated on the possibility of developing an artificial intelligence psychotherapy system and tutorial system.

This survey introduces the reader to that branch of artificial intelligence known as parallel distributed processing, or colloquially, neural networking. The computers with which we are familiar involve serial computing systems. At great speed they perform single operations in a central processing unit and thereby lies both their strength and their weakness. Their computational power is phenomenal, but is constrained by the time taken to transport data to and from the processing unit (the von Neumann trap). Since the early 1940s attempts have been made to model neurones, latterly with electronic components, with the idea of building a machine with many processors instead of one powerful processor, hence the concept of parallel processing. We will return to dedicated hardware later because initially it has proved possible and helpful to simulate parallel processing on existing digital processors.

From the start, conscious effort was made to model the computer activity on what was known of the physiology of cerebral neuronal functioning. In 1949 Donald Hebb had postulated that brain cells work not individually but in "assemblies". He claimed that brain representation was achieved by synaptic activity, and postulated that the more electrical activity that occurred at a synapse for a given input the more easily that synapse would fire in the future. The result was the creation of assemblies or clouds of cells which fired in response to given inputs. However, there was more to learning than simply the formation of associated electrocortical activity patterns. These associations would account for memory but additionally we have to account for perception, emotion and the ability to react appropriately to the inputting of data which has been previously unknown ("self-organisation"). With these ambitious goals in mind, the first network to be designed was a simple one come to be known as the "perceptron". In this model a series of input units were allocated arbitrary weightings and the resulting products were summated to represent an output. This output was measured against the target output and if the output was correct or that which was expected, then no action was taken. If the output was incorrect then a fixed fraction of each input was subtracted from its corresponding weight if the output had been too large, and vice versa if it had been too small. This fixed fraction was predetermined by the programmer and influenced the rate at which the programmer wished the network to adjust towards the desired solution. It was called the "learning rate" of the network. Such a simple perceptron was able to do certain tasks in pattern classification but was not able to tackle a pattern of the logical complexity of the "exclusive-or" function. That is to say, it was limited to linearly separable classes.

The next development to consider was the addition of a third or hidden layer of computing nodes of "neurones" so that we now had an input layer, a hidden layer, and an output layer, with each layer being connected to all the nodes in each adjacent layer. What we had to do was compare the output from our output layer to our expected output, work out an error value, transfer that back by adjusting the weightings between the output layer and the hidden layer and thence work backwards to the strength of connections between the input layer and the hidden layer. It transpired that using the analogy of the physiological concept of summation was helpful in this mathematical application. A neurone received incoming exitation through its boutons terminaux, but many boutons could be firing without the neurone firing. However, once a critical number of boutons was firing, then the addition of a relatively small number would cause the nerve cell itself to fire and, conversely, after that critical number had been passed, the addition of fairly large numbers of firing boutons would have no effect upon the firing phenomenon. This could be represented mathematically by a sigmoid or S-shaped function and it was found that the incorporation of such a sigmoid function in the maths allowed classes to be isolated which were not linearly separable. Furthermore, when we were calculating how much we should change our weightings, we required to take into account where on that sigma point the error amount lay, since the error amount would have a larger effect if it was on the steep exitation curve and a smaller effect if it was on the tails of the sigma. Therefore, the sigmoid function was built into the error estimation as well as the output function. This whole process of calculation of error and retrospective adjustment of the exitation states was called Back Error Propogation and led to the creation of networks which have important and surprising functional properties.

The fundamental property of the perceptron was that of pattern recognition; that is to say a system like this could be trained to classify patterns of input data such that it would classify data satisfactorily although it had never seen such data before. The process was similar to a straightforward psychological training exercise. The network was created, random weights were allocated and data sets were presented at the inputs in suitable form. The outputs were measured against the expected outputs and the network was informed of the discrepancies. Many sets were presented many times and as the iterations increased the error rate of the outputs diminished until a pre-arranged criterion was reached. At this point the network was in a stable state and could be said to be trained. It was now ready to receive data which it had never had presented before in order to test whether or not it was able to recognise the patterns inherent in this data. Of course, if data were presented which contained no patterns then such a network would be unable to detect patterns or classes. The addition of the hidden layer in the network allowed the quality of feature detection to emerge. Given varied inputs and trained on fixed outputs the network "learnt" to detect these features in the inputs which could identify the outputs. The analogy with diagnosis was close and such systems have been used to input symptoms and signs and output dermatological diagnoses (Yoon *et al*, 1989). If the hidden layer had many fewer nodes than the input layer then after training, the hidden layer contained a condensed representation of a great deal of data "distributed" across a few nodes and in effect held that data in condensed form (hence "parallel distributed processing").

Another type of network which may have applications for psychiatry was that known as the Hopfield Net. This network had one layer of nodes which were universally connected. Each element had a strength of plus one or minus one and each neurone updated itself periodically. The input to such a network was a string of numbers and the mathematical periodic updating of such a network was designed to converge towards ever lower values such that the system stabilised when no lower values could be obtained. This could result in the maths being trapped in a "local minimum" area, where in order to get to the lowest possible value a step upwards over higher values would be necessary, but this was mathematically impossible for the programme to do. It was therefore necessary to build in a function which could shake up the mathematical numbers to allow the procession towards the global minimum to be restarted. Such a system allowed the pairing of associated patterns and, in addition to this, if the data of a learned pair were presented in attenuated form then the network would output the corrected or total pattern, that is to say, it would recover a complete pattern given partial data. There is a tempting analogy here with that of the associative memory function.

We will now consider networks based on "competitive learning". These were two layer networks with each node connected to every node in the next layer. Once again the product of input numbers and random weights were computed, but in this case only the node with the highest value registered, the other values in the second layer of nodes remained unchanged. That is, there was one winner. The weights going to this winning unit were adjusted in order to make them more resemble the input pattern. In this way, without feeding the system correct solutions, the system identified patterns in the input data. The concept of a layer of nodes in competition with each other could be extended so that the "successful" node influenced the weightings in a positive direction while at the same time it influenced its immediate neighbours in a negative direction. This had the effect of maximising differences which the competitive function was isolating. Physiologically, we found the inhibition analogy at synaptic level, neuronal level and probably at cell assembly level.

There were other network models which used combinations of the principles discussed above. These included the Kohonen Feature Map which condensed data in a multidimensional form (c.f. the cerebral homunculus), and "Counter Propogation" which input a pattern, related this to its condensed representation of a state of patterns, and reproduced the nearest match (e.g. trace memories). Adaptive Resonance Theory allowed category formation without training on "correct" answers but had in addition a "novelty" detector which could be set high or low. If set high it would drive the network to create more rather than fewer categories, i.e. it scanned, as does the brain, for unfamiliar, or unexpected features in its environment.

We have drawn attention to analogies between brain functions and network functions where these seemed cogent. It would be naive to underestimate the failings in the analogy between networks and brain function. The astronomical numbers involved *in vivo* plus the existence of anatomical and biochemical subsystems made the most sophisticated network appear rudimentary. However, this very comparison emphasised the surprising extent of the success achieved in mimicking brain functions. What of the future? Already parallel processing computers exist and the problems of coordinating the activities of the individual processing units are being worked out. *The Times* (11 January 1990) has reported the development of a chip designed to receive multiple inputs and to output through a single channel. This would effectively produce an analogue system as opposed to the digital systems we have been describing above. We have seen how the functions of existing networks complement the computing power of digital computers and as an indicator of their assessed potential, *New Scientist* (6 January 1990) announced that Japan's Ministry of International Trade and Industry is planning to launch a national research project on neural computers to run for ten years and replace their Fifth Generation Computer Programme. On a more modest note, McDonald & McDonald (1990) and Lucas (1990) have announced possibly the first clinical uses of such networks in psychiatry. We foresee their use in diagnosis in symptom cluster recognition and in cognitive modelling, e.g. the dementia process could be subjected to a process of reverse engineering by reducing hidden nodes one by one or by eliminating input measures one by one and noting the effects on total performance. Here is a conceptual tool which allows us to model complicated cerebral functions and the profession would be unwise to spurn it as too simplistic or mechanical.

References

BODEN, M. (1977) *Artificial Intelligence and Natural Man.* Hemel Hempstead: Harvester Press.

HEBB, D. O. (1949) In *The Organization of Behaviour.* New York: Wiley.

LUCAS, C. P. (1990) Neural network technology. *Psychiatric Bulletin,* **14,** 237–238. (Correspondence).

MCDONALD, S. & MCDONALD, C. (1990) Clinical diagnosis by neural networking using psychometric data. *Psychiatric Bulletin,* **14,** 45–46. (Correspondence).

SERVAN-SCHREIBER, D. (1986) Artificial intelligence and psychiatry. *The Journal of Nervous and Mental Disease,* **174,** 191–202.

WEXLER, B. (1986) A model of brain function: its implications for psychiatric research. *British Journal of Psychiatry,* **148,** 357–362.

YOON, Y., BROBST, R. W., BERGSTRESSER, P. R. & PETERSON, L. (1989) A desktop network for dermatology diagnosis. *Journal of Neural Network Computing,* **1,** 43–52.