

A celebration of connectionism

New developments in neural network theory have excited both psychologists and neurobiologists. Practitioners of the new art displayed their wares last week.

Oxford

WHEN David Rumelhart, Geoffrey Hinton and Ronald Williams described for neural networks a powerful new learning procedure called back-propagation (*Nature* 323, 533; 1986), they noted that theirs was not a plausible model of how brains learn. Yet the generality of their approach, and the several intriguing features of network learning by back-propagation which have come to light, have stimulated a resurgence of interest in neural network models among neuroscientists, theoreticians and experimentalists alike. Last week, at a meeting organized by the Society of Experimental Psychology, a packed audience heard Geoffrey Hinton describe a new learning algorithm that seems a better model of biological learning than is back-propagation by parallel networks which nevertheless seems to retain much of the power of its predecessor.

In parallel distributed processing, a network can be thought of as embodying a mathematical function mapping vectors in 'input space' to vectors in 'output space', much as matrices effect linear transformations between vector spaces. A vector in the neural context is simply the pattern of excitation of some set of units taken to be the input or the output of the network. The processing is thus distributed in the pattern of the connections between units of the network and their strengths. Corresponding to the real physiological task of, say, pattern recognition, will be some kind of network function mapping inputs (patterns) onto outputs (interpretations). The all-important question is what kind of network is needed for a particular task.

Simple networks developed in the 1960s, known as perceptrons, involved only two layers of units, an input and an output, with direct connections between them. Such networks are very limited in the range of tasks they can carry out. The versatility of a network can be greatly increased by the introduction of intermediate layers of 'hidden' units, but this raises the problem of how it can be trained.

Back-propagation provides an elegant way of training a multi-layered network. During learning, the output vectors generated by the network for a given input are compared with the desired output, giving an error calculated from the difference between the two. Back-propagation calculates the dependence of this error on all the connection weights, simply by using

the chain-rule for differentiation, and the weights are adjusted to reduce the error, so that the network converges by gradient descent on the required structure.

During learning, the network comes to capture certain general features which are characteristic of its task. The hidden units, in particular, develop features that seem especially significant to neuroscientists who record the properties of single neurons in brains. For example, in some cases they are reminiscent of the way in which some neurons in the brain are found to be specific for different aspects of the representation of the visual field.

Even so, this system of learning by back-propagation has not seemed very biologically realistic. Hinton (Carnegie-Mellon University) and his colleagues have been looking for a more plausible system of learning.

The new development is known as a 'recirculation' algorithm, and works as follows. In a network learning by back-propagation, there is a linear flow of activity (via the hidden intermediate units) from the input units to the output units. In the new system, the hidden units connect back to the single layer of 'visible' input units. Activity thus recirculates through the network; during training, the connection weights are adjusted to minimize the rate of change of activity in each unit. Thus, when trained, the network is set up so as to stabilize on certain states, and so can work as a kind of 'content addressable memory' with the property that degraded or incomplete forms of the training inputs can regenerate the correct version.

It has been shown that, under certain conditions, the new algorithm is equivalent to gradient descent, and it has been found empirically that the system still works when these conditions are relaxed.

A number of interesting applications of back-propagation were reported at the meeting. Hinton described a network for recognizing one-dimensional shapes on a one-dimensional retina independently of position: the hidden units learn to respond to shapes in different positions. Hinton also described a speech-recognition network which learns to recognize spoken consonants given very noisy corrupted data. It appears to perform almost as well as people, and better than the previous best system of automated speech recognition.

Several speakers described analogies between the behaviour of their networks during training or after 'damage' and what

is known of human learning and cognitive disorders. For example, J. L. McClelland (Carnegie-Mellon University) described how a network for learning a balance-beam task progressed through stages of competence similar to those of children given the same task. The problem is to decide which way a balance-beam will tip, depending on the position and size of weights on either side of the fulcrum. With an appropriately biased learning environment, such as children might well experience, the network, like children, initially bases its decisions purely on weight information; gradually the network learns to use the position of the weights. M.S. Seidenberg (McGill University) described a network for word recognition and pronunciation that, when 'damaged' by the removal of hidden units, displayed behaviour reminiscent of some human disorders, such as dyslexia.

Parallel distributed processing is not without its critics, and S. Pinker (MIT) reported that a linguistic analysis of Rumelhart and McClelland's network for changing the tense of verbs in sentences shows that the system is not 'descriptively adequate' as a model for human language, in that it abandons certain symbolic rules and principles that linguistic studies suggest are crucial in human language.

David Willshaw (Edinburgh) asked whether parallel distributed processing networks might, like perceptrons, similarly cease to make significant progress and fade in interest after a period of development and excitement. McClelland's riposte was that work on perceptrons was severely limited by the available computer power and circumstances are now sufficiently different to justify optimism.

The biological relevance of parallel distributed processing remains an open question. Independently of relevance, however, work on network systems may be of interest at a purely theoretical level. The present work is a kind of experimental mathematics, and in that respect is rather similar to that of Mandelbrot on fractals, also made possible and inspired by computers. The hope is that, in future, deductive proofs will give a more rigorous basis to work on networks.

Many interesting problems remain. On what set of functions will a given network topology converge? How can the optimal network for a given task be predicted, and how long will training take? And so on.

Geoffrey North