FORUM ON CONNECTIONISM

Language Learning in Massively-Parallel Networks

Terrence J. Sejnowski
Biophysics Department
Johns Hopkins University
Baltimore, MD 21218

PANELIST STATEMENT

Massively-parallel connectionist networks have traditionally been applied to constraint-satisfaction in early visual processing (Ballard, Hinton & Sejnowski, 1983), but are now being applied to problems ranging from the Traveling-Salesman Problem to language acquisition (Rumelhart & McClelland, 1986). In these networks, knowledge is represented by the distributed pattern of activity in a large number of relatively simple neuron-like processing units, and computation is performed in parallel by the use of connections between the units.

A network model can be "programmed" by specifying the strengths of the connections, or weights, on all the links between the processing units. In vision, it is sometimes possible to design networks from a task analysis of the problem, aided by the homogeneity of the domain. For example, Sejnowski & Hinton (1986) designed a network that can separate figure from ground for shapes with incomplete bounding contours. Constructing a network is much more difficult in an inhomogeneous domain like natural language. This problem has been partially overcome by the discovery of powerful learning algorithms that allow the strengths of connection in a network to be shaped by experience; that is, a good set of weights can be found to solve a problem given only examples of typical inputs and the desired outputs (Sejnowski, Kienker & Hinton, 1986; Rumelhart, Hinton & Williams, 1986).

Network learning will be demonstrated for the problem of converting unrestricted English text to phonemes. NETtalk is a network of 309 processing units connected by 18,629 weights (Sejnowski & Rosenberg, 1986). It was trained on the 1,000 most common words in English taken from the Brown corpus and achieved 98% accuracy. The same network was then tested for generalization on a 20,000 word dictionary: without further training it was 80% accurate and reached 92% with additional training. The network mastered different letter-to-sound correspondence rules in varying lengths of time; for example, the "hard c rule", c -> /k/, was learned much faster than the "soft c rule", c -> /s/.

NETtalk demonstrably learns the regular patterns of English pronunciation and also copes with the problem of irregularity in the corpus. Irregular words are learned not by creating a look-up table of exceptions, as is common in commercial text-to-speech systems such as DECtalk, but by pattern recognition. As a consequence, exceptional words are incorporated into the network as easily as words with a regular pronunciation. NETtalk is being used as a research tool to study phonology; it can also be used as a model for studying acquired dyslexia and recovery from brain damage; several interesting phenomena in human learning and memory such as the power law for practice and the spacing effect are inherent properties of the distributed form of knowledge representation used by NETtalk (Rosenberg & Sejnowski, 1986).

NETtalk has no access to syntactic or semantic information and cannot, for example, disambiguate the two pronunciations of "read". Grammatical analysis requires longer range interactions at the level of word representations. However, it may be possible to train larger and more sophisticated networks on problems in these domains and incorporate them into a system of networks that form a highly modularized and distributed language analyzer. At present there is no way to assess the computational complexity of these tasks for network models; the experience with NETtalk suggests that conventional measures of complexity derived from rule-based models of language are not accurate indicators.

REFERENCES

Ballard, D. H., Hinton, G. E., & Sejnowski, T. J., 1983. Parallel visual computation, **Nature** 306: 21-26.

Rosenberg, C. R. & Sejnowski, T. J. 1986. The effects of distributed vs massed practice on NETtalk, a massively-parallel network that learns to read aloud, (submitted for publication).

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. 1986. In: **Parallel Distributed Processing: Explorations in the Microstructure of Cognition.** Edited by Rumelhart, D. E. & McClelland, J. L. (Cambridge: MIT Press.)

Rumelhart, D. E. & McClelland, J. L. (Eds.) 1986. **Parallel Distributed Processing: Explorations in the Microstructure of Cognition.** (Cambridge: MIT Press.)

Sejnowski, T. J., Kienker, P. K. & Hinton, G. E. (in press) Learning symmetry groups with hidden units: Beyond the perceptron, **Physica D.**

Sejnowski, T. J. & Hinton, G. E. 1986. Separating figure from ground with a Boltzmann Machine, In: **Vision, Brain & Cooperative Computation,** Edited by M. A. Arbib & A. R. Hanson (Cambridge: MIT Press).

Sejnowski, T. J. & Rosenberg, C. R. 1986. NETtalk: A parallel network that learns to read aloud, Johns Hopkins University Department of Electrical Engineering and Computer Science Technical Report 86/01.