-n

MODULARITY OF MIND, CEREBRAL LOCALISATION AND CONNECTIONIST NEUROPSYCHOLOGY Book Review of Miikkulainen on Language-Network

-n
Robert William Kentridge
Psychology Department
Univerity of Durham
Durham DH1 3LE England

robert.kentridge@durham.ac.uk

Abstract

Although Miikkulainen's DISCERN (1993) is a model of cognition and not the brain, its completely connectionist construction and similarities between the effects of damage on its behaviour and the neuropsychology of dyslexia may tempt us to treat it as a brain model. In particular, we may draw parallels between the architecture of DISCERN and localisation of function in the cortex. I argue that "modularity" of the brain may be organised in terms of time-scales as well as space. Forms of "brain-damage" in connectionist neuropsychology may differ profoundly from those causing similar effects in people.

Keywords

computational modeling, connectionism, distributed neural networks, episodic memory, lexicon, natural language processing, scripts.

I. INTRODUCTION

1. Miikkulainen (1993) presents a clear and complete picture of his parallel distributed processing model of natural language processing, DISCERN. His aim in developing the model was to show that a complete natural language processing system could be built entirely from artificial neural network components; however, his work also raises interesting questions about the implementation of natural language processing in real neural networks: brains. Two design decisions make his work pertinent to biological questions: first, the decision to exclude any explicitly symbolic processing from the system, and second, the decision to model the integration of many components of natural language processing rather than concentrating on an individual subprocess. The consequence of these decisions is that DISCERN is a modular parallel system. The brain is obviously fundamentally parallel, and neuropsychological and neuroanatomical evidence suggest that both its function and its organisation are modular (e.g., Fellerman & Van Essen, 1991; Ellis & Young, 1988). To what extent can parallels be drawn between modularity of mind as exemplified by DISCERN and cerebral localisation of function which might be revealed by localised brain damage?

2. Before addressing this question I must emphasise that Miikkulainen does not suggest that DISCERN is a brain model. My objective is to discuss how neuroscientists might interpret models such as DISCERN; it is not to criticise DISCERN which is clearly a successful piece of connectionist cognitive modelling.

II. COMPUTATIONAL FUNCTION AND DYNAMICS

3. We can broadly categorise artificial neural network modules in terms of the dynamics underlying their information processing. Modules without dynamics, that is, networks in which stimuli are directly mapped onto outputs, have different capabilities from those governed by attractor dynamics, which are, in turn, different from those with still more complex dynamics extending over time. In attractor networks, the dynamics of the net cause information to be discarded from stimuli as sets of initial points corresponding to those stimuli converge onto an attractor. We can view this process as labelling all points in a basin of attraction with a single symbol. The relationship between different basins of attraction and stimuli determines the response of a network to new stimuli. In an attractor network where stimuli do not modify the activities of all units, the response to a stimulus depends on which attractor the network has converged to prior to the introduction of the next stimulus and the nature of the stimulus itself. In formal computational terms, taking basins of attraction to correspond to the states of a formal machine, an attractor network has the computational power of a finite-state automaton. That is, transitions between states in the machine depend only on the current state of the machine and the input presented to it. In these terms, feedforward networks which simply perform mappings are less powerful. Although the power of computation involved in training them is considerable, their response to stimuli does not depend at all on the prior stimuli they encountered. Networks which perform more powerful computation in the course of responding to a stimulus must have behaviour which depends on sequences of prior stimuli, rather than simply the input and immediately preceding stimulus. Complete convergence onto an attractor will terminate this sequence, so the ideal network's response to a stimulus will be an infinite transient whose path is governed by the input and the attractor structure of the network.

4. In DISCERN the response of feature maps to stimuli is a based on simple mappings (although adapting those maps involves more complex dynamics mediated through lateral connections; these lateral connections do not play a part in the response of the network). The episodic trace-memory modification to feature maps involves the addition of attractor dynamics to the feature maps during their response to stimuli. The FGREP modules responsible for the initial encoding of stimuli are purely feedforward in their initial incarnation, but the modified recurrent FGREP modules can, in principle, exhibit complex dynamics supporting expectations based on sequences of previously encountered stimuli. There is evidence that both simple mappings and attractor dynamics are involved in natural language processing; as Miikkulainen notes, both DISCERN and the Plaut & Shallice (1993) connectionist model of acquired dyslexia mimic different categories of dyslexic errors when the attractor and mapping components of the models are damaged. If a model of natural language processing is to cope with complex syntactic structures (which DISCERN does not attempt to) then it must be able to detect structure within sequential input, requiring more complex dynamics, similar to those supported by DISCERN's recurrent FGREP

modules.

5. What features of artificial networks produce these different dynamics? Obviously, different training algorithms are used to produce different behaviours; there are, however, more important fundamental architectural constraints. Feedforward networks must, as their name implies, consist only of connections between neurons operating in a single direction through the network -- their connection pattern must be acyclic. Networks which contain cyclic connections, but in which all connections between pairs of neurons are symmetric, will converge to point attractors. When this symmetry constraint is relaxed, more complex behaviour is possible, although its precise nature depends on additional factors such as levels of background noise and constraints on the range of connections.

III. DYNAMICS AND ANATOMY

6. In DISCERN, modules having different functions for the most part have different architectures. This separation of modules by function and "anatomy" is appealing, because it may explain correspondences between deficits in the model's performance when specific modules are damaged and deficits found in people which might be attributed to localised brain damage. Indeed, Miikkulainen discusses the importance of "connectionist neuropsychology" in the future understanding of representation and processing in connectionist models, both in its own right and as a tool for understanding neuropathologies. Superficially, the problem with this is that the different areas of the cortex involved in natural language processing do not differ fundamentally in their architecture. One of the remarkable things about the cortex is its uniformity. Nevertheless, the neuropsychological evidence for modularity, combined with convincing connectionist models such as DISCERN and that of Plaut & Shallice (1993), suggests that the cortex is performing a number of types of computations which brain damage can effect differentially.

7. How can this be? I will discuss a number of possibilities. My aim is not to choose one or another of them; rather, it is to point out that the separation of computational function in the cortex may depend on factors not immediately apparent in a connectionist model of neuropsychology. The bulk of cortico-cortical connections can be divided into two classes (Braitenberg & Schuz, 1991). First, there are those connections that are mediated by the basal dendritic fields of pyramidal cells in which the probability of a synapse occurring between two cells varies inversely with the distance between them. Second, there are those that ar mediated by the apical dendritic fields of pyramidal cells in which the probability of a synapse occurring between two cells is independent of their spatial separation. Note, however, that neither of these subdivisions provides us with either a feedforward system or a symmetrically interconnected one. The long range apical system, however, is organised in its projections between cortical areas, although not in a space-dependent fashion. If one constrains the time-scale over which interactions occur to be short, it might be reasonable to regard these long range connections as forming a number of distinct feedforward systems. Any signals projected back might be assumed to take too long to arrive to modify the initial feedforward mapping of a stimulus.

8. It is more difficult to resolve the differences between the short range connection system and the requirements of DISCERN. If one pursues the notion that trace memory requires attractor dynamics, one possibility is to regard spatially localised COLLECTIONS of cells as the components of a system. If we assume some randomness in the strength of individual connections, then, on average, the strength of connections between groups of cells will be approximately symmetric over a short range. This approximation to symmetry becomes weaker at longer ranges, where there are fewer connections between groups to sum over. Eventually, these long-range asymmetries may destabilise "short-range" attractors. Unfortunately, this scheme cannot be applied at the single cell level. The argument relies on there being large numbers of connections between groups of cells at short range. Although the probability of connection between individual cells is space- dependent, this does not imply multiple connections between pairs of nearby cells. In fact, very few pyramidal cells make multiple synaptic contacts (Braitenberg & Schuz, 1991). It therefore becomes problematic to assume that groups of neurons will behave as a unit over long time-scales. As intrinsically unstable units, their behaviour is subject to change independent of their interactions with other neuronal groups. This contrasts with improvement in the

approximation to symmetric intergroup connections as more individual connections are activated with time. If this "group approximation" to a symmetrically connected attractor net is plausible, it too is highly constrained in the time-scale over which it can operate effectively.

9. An alternative possibility for the implementation of trace-memory is that networks of neurons with spatially localised connections behave as selective dynamic patterns, forming systems akin to reaction-diffusion systems or excitable media (Gaponov-Grekhov & Rabinovich, 1992). If this is the case, then factors such as the time and space constants for the decay of activity (which may depend on the relative strengths and distributions of excitatory and inhibitory neurons) and the level of "background" activity in the system becomes crucial.

IV. CONCLUSIONS

10. In these speculations on possible neural realisations of some of the functions required by DISCERN, I have tried to suggest how functions may be separated by time-scales rather than anatomy. The candidates for specific control of these systems, which may explain neuropsychological dissociations of function, are just as likely to be neurochemical systems operating over different time-courses as anatomical localisations. My point is not to suggest that Miikkulainen has been driven by such assumptions (although he does speculate about the biological implementation of DISCERN). As cognitive models such as DISCERN become fully connectionist, however, it is tempting to view them directly as brain models. Sometimes it may be appropriate to equate "lesions" of connectionist models with localised brain damage; in other cases, however, brain dysfunction may be attributable to damage to remote or diffuse modulatory systems. Great care is therefore required in drawing inferences from neuropsychology to support modular connectionist models of cognition or using these models to explain the basis of neuropatholgies.

11. Models such as DISCERN clearly add to our understanding of cognitive processes and may inform the production of brain models, although in themselves they are not brain models. I may be tackling a straw man by suggesting that DISCERN and its descendants are in danger of being viewed as brain models, but I hope that examining how their components might be implemented in brains increases their usefulness across a range of disciplines.

REFERENCES

Braitenberg, V. & Schuz, A. (1991) Anatomy of cortex: Statistics and geometry. Berlin: Springer-Verlag.

Ellis, A.W. & Young, A.W. (1988) Human cognitive neuropsychology. Hove: Lawrence Erlbaum.

Fellerman, D.J. & Van Essen, D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex. Cortex ,1:1-47.

Gaponov-Grekhov, A.V. & Rabinovich, M.I. (1992) Nonlinearities in action: Oscillations, chaos, order, fractals. Berlin: Springer-Verlag.

Miikkulainen, R. (1993) Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory. Cambridge MA: MIT.

Miikkulainen, R. (1994) Precis of: Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon and Memory. PSYCOLOQUY 5(46) language-network.1.miikkulainen.

Plaut, D.C & Shallice, T. (1993) Deep dyslexia: A case study of connectionist neuropsychology. Cognitive Neuropsychology, 10:377-500.

Topic: language network [View Topic]

Article: 9) kentridge 5(87) MODULARITY OF MIND, CEREBRAL [View Article]

Topic: language network [View Topic]

Article: 9) kentridge 5(87) MODULARITY OF MIND, CEREBRAL [View Article]