An Experimental Test of Rule-like Network Performance

Maartje E. J. Raijmakers & P.C.M. Molenaar
Dept. of Psychology, University of Amsterdam, Roetersstr 15, 1018 WB Amsterdam, The Netherlands
op_raijmakers@macmail.psy.uva.nl

Introduction

To what extent can the behavior of a neural network performing a classification task, for example the balance scale, be called rule based? On the one hand, obviously, explicit rules are not available, although the responses of a network might be consistent with rules. On the other hand, networks are not thought to learn fully in accordance with behaviorist theories and are believed to have particular cognitive processing properties such as selective encoding of input patterns. The hidden units should function as mediating concepts. However, it is impossible to study the internal state of the human brain in the same detail as the internal structure of a neural network. We propose a simple empirical test for neural networks that discriminates between the formation of stimulus-driven associations and the formation of cognitive concepts: the discrimination-shift task (Kendler, 1995).

Discrimination-Shift Task

In the discrimination-shift task subjects learn to discriminate on the basis of reinforcement contingencies between four stimuli which are presented in two distinct pairs. The stimuli are distinguishable on two dimensions: for example, shape (round/triangle) and color (white/black). Each stimulus pair appears in two configurations of which only the positions of the stimuli differ. The experiments start with a pre-shift phase during which an initial discrimination of the stimuli is learned. The pre-shift phase continues until the number of correct responses in a sequence of adjacent trials meets a given criterion. The pre-shift is followed by either a reversal shift (RS) or an extradimensional shift (EDS). In both cases, the reinforcement is changed without informing the subject. A RS implies that all stimuli that received positive reinforcement get negative reinforcement, and vice versa. An EDS means that the dimension upon which the reinforcement is based, shape or color, is shifted. According to, for example, Kendler (1995) animals learn a simple discrimination task by forming simple stimulus-driven associations, since they learn an EDS faster than a RS, and humans (older than 6 years) learn the same task by forming mediated concepts, since they learn a RS faster than an EDS.

We applied the test and related tasks to a PDP-model that is previously used to simulate the acquisition of increasingly complex rules on the balance scale task (McClelland & Jenkins, 1991). We varied the number of hidden units and the connection structure between input and hidden units, being constraint, C, or unconstrained, UC. Most tested network architectures learn an EDS faster than a RS. This accords with behaviorist models. The 8-2-2 C networks show no difference between the number of learning cycles in the RS phase and the EDS phase. This gives no clear indication for the learning mode of this network. Therefore, the 8-2-2 C networks are examined by means of a trial-by-trial analysis. The result of this analysis agrees fully with predictions from behaviorist theories and results of 4-year-old children.

Optional-Shift Task

A second experimental design is the optional-shift task (Kendler, 1995). The optional shift task starts with the same pre-shift phase as the first described task. During the phase that follows, the shift-discrimination phase, only one of the two stimulus pairs is learned with reversed reinforcement, such that the reinforcement of this stimulus pair agrees with both shifts: a RS and an EDS. During the test series, that is, after the attainment of criterion in the shift-discrimination phase, all stimulus pairs are presented again, but without reinforcement. The responses on these test trials indicate a RS, an EDS, are both right, or are both left. Also in this task, the behavior of networks of all tested configurations respond most of the time with an EDS. Only in a few cases, a RS is performed. If the number of learning trials of the networks is increased, that is, the criterion for an output node being active or inactive is more severe, the responses agree more often with an EDS. These results agree with behavior found in rats but not with children.

Conclusion and Discussion

Simulations show that the learning behavior of all tested network configurations is equivalent to forming stimulus-driven associations, which agrees with behaviorist models and the discrimination-shift behavior of rats. The question now is which alternative architectures will lead to discrimination-shift behavior that coincides with human subjects: children or adults. One aspect of the simulation that could be changed is the representation of the stimuli. Almost no feature extraction takes place, as in the balance scale simulation study. Furthermore, stimuli characteristics of completely different nature, such as color and form, are processed by one module (in the unconstrained networks) or equivalent modules (in the constraint network). A third possible change is adding a sort of attention and bias to a network (see also Kruschke, 1992).

References

Kendler, T.S. (1995). *Levels of Cognitive Development.* Mahwah, New Jersey: Lawr.Erlb.Ass.

Kruschke, J.K. (1992). ALCOVE: An Exemplar-Based Connectionist Model of Category Learning. Psy. Rev., 99(1).

McClelland, J.L., & Jenkins, E. (1991). Nature, nurture, and connections: implications of connectionist models for cognitive development. In (Ed.), *Architectures for intelligence: The twenty-second Carnegie Mellon Symposium on cognition.* (pp. 41-73). Pittsburgh: Lawr. Erlb. Ass.

Raijmakers, M.E.J., van Koten, S., & Molenaar, P.C.M. (1996). On the validity of simulating stagewise development by means of PDP networks: Application of Catastrophe Analysis and an experimental test of rule-like network performance. Cognitive Science, 20 (1).