A Neural Network Model of Discrimination Shifts

Sylvain Sirois and **Thomas R. Shultz**
Department of Psychology
McGill University
1205 Penfield Avenue. Montreal, Canada. H3A 1B1
{sirois,shultz}@psych.mcgill.ca

Introduction

It was recently claimed that feedforward neural networks cannot simulate what is considered to be rule-based behavior in adult humans (Raijmakers, van Koten, & Molenaar, 1996). Back-propagation networks learning discrimination shifts were found to behave in an associative way characteristic of young children. We suggest that these simulations are flawed by using layered networks on linear problems.

Discrimination shift tasks

Discrimination shift paradigms involve the pairwise discrimination of stimuli with attributes on three binary dimensions. A pair of stimuli exhibit mutually exclusive attributes on all three dimensions, a constraint that allows four pairs. Participants learn to identify the stimulus in each pair that exhibits the targeted attribute (e.g., dark). Learning occurs by reinforcement until participants reliably identify the target. When criterion is reached, shifts in reward contingencies can be introduced.

In a reversal shift (RS), training is shifted to the other attribute of the same dimension (e.g., from dark to bright). All reinforcement contingencies are thus changed. For extra-dimensional shifts (EDS), training is shifted to an attribute of a previously irrelevant dimension (e.g., dark to small). Contingencies are changed for only half of the pairs (e.g., half of the dark stimuli are also small). Human adults typically learn an RS faster than an EDS and their performance on unchanged pairs of an EDS is impaired during shift learning (Kendler, 1983; Tighe & Tighe, 1978).

This has suggested that adults learn these tasks by way of mediating concepts representing the relevant dimension (Kendler, 1983). But when younger children lacking mediating ability are trained for an extended time beyond criterion, they too perform an RS faster than an EDS (Wolff, 1967). This is the overtraining effect, and it can be implemented in neural networks by lowering the score threshold producing deeper learning.

We propose a model of discrimination shifts using cascade-correlation networks. We trained 120 networks on RS and EDS tasks using a small score threshold. The networks learned an RS faster than an EDS and showed impaired performance on unchanged EDS pairs during shift training. They did not recruit hidden units.

In an optional shift (OS), initial training is identical to that in RS and EDS, but at the onset of the shift, two of the four pairs are presented with changed reward contingencies such that the shift agrees with either an RS or an EDS. For most adults, responses on the test pairs agree with an RS. This is also the case of young children in overtraining conditions (Wolff, 1967). We trained 60 networks on an OS task, again with a small score-threshold. The behavior of 57 networks was consistent with an RS, a proportion not significantly different from that observed in adults. Again, hidden units were not recruited by the networks.

Discussion

Our networks showed behavior consistent with what is found in adults on RS, EDS, and OS tasks. Moreover, the lack of hidden units in the networks implies that mediated processing is not involved in their behavior. We suggest that human adults may submit themselves to spontaneous overtraining through a process similar to rehearsal in memory tasks. This has already been suggested by Levine (1975) for discriminative learning. We also modeled the learning of young children using a higher score-threshold (Sirois & Shultz, in preparation). Our model shows that PDP networks can simulate behavior usually considered rule-based. Indeed, the prospects look favorable for simulating a wide range of age-related phenomena in the discrimination shift literature.

References

Kendler, T.S. (1983). Labeling, overtraining and levels of function. In T.J. Tighe & B.E. Shepp (Ed.), *Perception, Cognition, and Development: Interactional Analysis* (pp. 129-162). Hillsdale, NJ: Lawrence Erlbaum Associates.

Levine, M. (1975). *A cognitive theory of learning: research on hypothesis testing*. Hillsdale, NJ: Erlbaum.

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*, 101-136.

Sirois, S., & Shultz, T.R. (in preparation). *Neural Network Models of Discrimination Shifts: A Developmental Approach.*

Tighe, T.J., & Tighe, L.S. (1978). A perceptual view of conceptual development. In R.D. Walk & H.L. Pick, Jr (Eds.), *Perception and Experience* (pp. 387-416). New York, NY: Plenum Press.

Wolff, J.L. (1967). Concept-shift and discrimination-reversal learning in humans. *Psychological Bulletin, 68*, 369-408.