A

Artificial Neural Networks and Cognitive Modelling

Amanda J.C. Sharkey
University of Sheffield, UK

INTRODUCTION

In their heyday, artificial neural networks promised a radically new approach to cognitive modelling. The connectionist approach spawned a number of influential, and controversial, cognitive models. In this article, we consider the main characteristics of the approach, look at the factors leading to its enthusiastic adoption, and discuss the extent to which it differs from earlier computational models. Connectionist cognitive models have made a significant impact on the study of mind. However connectionism is no longer in its prime. Possible reasons for the diminution in its popularity will be identified, together with an attempt to identify its likely future.

The rise of connectionist models dates from the publication in 1986 by Rumelhart and McClelland, of an edited work containing a collection of connectionist models of cognition, each trained by exposure to samples of the required tasks. These volumes set the agenda for connectionist cognitive modellers and offered a methodology that subsequently became the standard. Connectionist cognitive models have since been produced in domains including memory retrieval and category formation, and (in language) phoneme recognition, word recognition, speech perception, acquired dyslexia, language acquisition, and (in vision) edge detection, object and shape recognition. More than twenty years later the impact of this work is still apparent.

BACKGROUND

Seidenberg and McClelland's (1989) model of word pronunciation is a well-known connectionist example. They used backpropagation to train a three-layer network to map an orthographic representation of words and non-words onto a distributed phonological representation, and an orthographic output representation. The model is claimed to provide a good fit to experimental data from human subjects. Humans can make rapid decisions about whether a string of letters is a word or not, (in a lexical decision task), and can readily pronounce both words and non-words. The time they take to do both is affected by a number of factors, including the frequency with which words occur in language, and the regularity of their spelling. The trained artificial neural network outputs both a phonological and an orthographic representation of its input. The phonological representation is taken as the equivalent to pronouncing the word or non-word. The orthographic representation, and the extent to which it duplicates the original input, is taken to be the equivalent of the lexical decision task

The past tense model (McClelland & Rumelhart, 1986) has also been very influential. The model mirrors several aspects of human learning of verb endings. It was trained on examples of the root form of the word as input, and of the past-tense form as output. Each input and output was represented as a set of context-sensitive phonological features, coded and decoded by means of a fixed encoder/decoder network. A goal of the model was to simulate the stage-like sequences of past tense learning shown by humans. Young children first correctly learn the past tense of a few verbs, both regular (e.g. looked) and irregular (e.g. went, or came). In stage 2 they often behave as though they have inferred a general rule for creating the past tense, (adding –ed to the verb stem). But they often over-generalise this rule, and add –ed to irregular verbs (e.g comed). There is a gradual transition to the final stage in which they learn to produce the correct past tense form of both regular and exception words. Thus their performance exhibits a U-shaped function for irregular verbs (initially correct, then often wrong, then correct again).

The model was trained in stages on 506 English verbs. First, it was trained on 10 high frequency verbs (regular, and irregular). Then medium frequency verbs (mostly regular) were introduced and trained for a number of epochs. A dip in performance on the irregular verbs occurred shortly after the introduction of the medium frequency verbs – a dip followed by a gradual improvement that resembled the U-shaped curve found in human performance.

THE STRENGTHS AND LIMITATIONS OF CONNECTIONIST COGNITIVE MODELLING

The models outlined above exhibit five typical features of connectionist models of cognition: (i) They provide an account that is related to and inspired by the operations of the brain; (ii) They can be used both to model mental processes, and to simulate the actual behaviour involved; (iii) They can provide a 'good fit' to the data from psychology experiments; (iv) The model, and its fit to the data, is achieved without explicit programming and (v) They often provide new accounts of the data. We discuss these features in turn.

First there is the idea that a connectionist cognitive model is inspired by, and related to, the way in which brains work. Connectionism is based on both the alleged operation of the nervous system and on distributed computation. Neuron-like units are connected by means of weighted links, in a manner that resembles the synaptic connections between neurons in the brain. These weighted links capture the knowledge of the system; they may be arrived at either analytically or by "training" the system with repeated presentations of input-output training examples. Much of the interest in connectionist models of cognition was that they offered a new account of the way in which knowledge was represented in the brain. For instance, the behaviour of the past tense learning model can be described in terms of rule following – but its underlying mechanism does not contain any explicit rules. Knowledge about the formation of the past tense is distributed across the weights in the network.

Interest in brain-like computing was fuelled by a growing dissatisfaction with the classical symbolic processing approach to modelling mind and its relationship to the brain. Even though theories of symbol manipulation could account for many aspects of human cognition, there was concern about how such symbols might be learnt and represented in the brain. Functionalism (Putnam, 1975) explicitly insisted that details about how intelligence and reasoning were actually implemented were irrelevant. Concern about the manipulation of meaningless, ungrounded symbols is exemplified by Searle's Chinese Room thought-experiment (1980). Connectionism, by contrast, offered an approach that was based on learning, made little use of symbols, and was related to the way in which the brain worked. Arguably, one of the main contributions that connectionism has made to the study and understanding of mind has been the development of a shared vocabulary between those interested in cognition, and those interested in studying the brain.

The second and third features relate to the way in which artificial neural nets can both provide a model of a cognitive process and simulate a task, and provide a good fit to the empirical data. In Cognitive Psychology, the emphasis had been on building models that could account for the empirical results from human subjects, but which did not incorporate simulations of experimental tasks. Alternatively, in Artificial Intelligence, models were developed that performed tasks in ways that resembled human behaviour, but which took little account of detailed psychological evidence. However, as in the two models described here, connectionist models both simulated the performance of the human tasks, and were able to fit the data from psychological investigations.

The fourth feature is that of achieving the model and the fit to the data without explicit handwiring. It can be favourably contrasted to the symbolic programming methodology of Artificial Intelligence, where the model is programmed step by step, leaving room for ad hoc modifications and kludges. The fifth characteristic is the possibility of providing a novel explanation of the data. In their model of word pronunciation, Seidenberg and McClelland showed that their artificial neural network provided an integrated (single mechanism) account of data on both regular and exception words where previously the old cognitive modelling conventions had forced an explanation in terms of a dual route. Similarly, the past-tense model was formulated as a challenge to rule-based accounts: although children's performance can be described in terms of rules, it was claimed that the model showed that the same behaviour could be accounted for by means of an underlying mechanism that does not use explicit rules.

In its glory days, connectionism's claims about novel explanations of stimulated much debate. There was also much discussion of the extent to which connectionism could provide an adequate account of higher mental processes. Fodor and Pylyshyn (1988)

Related Content