A Model of Innately Guided Learning by a Neural Network: The Case of Featural Representation of Speech

Ramin Charles Nakisa
Department of Experimental Psychology, Oxford University
Oxford OX1 3UD
ramin@psy.ox.ac.uk

Introduction

Newly born infants are able to finely discriminate almost all human speech contrasts and their phonemic category boundaries are initially identical, even for phonemes outside their target language. A connectionist model of innately guided learning is described which accounts for this ability. The approach taken has been to develop a model of innately guided learning in which an artificial neural network (ANN) is stored in a "genome" which encodes its architecture and learning rules. The space of possible ANNs is searched with a genetic algorithm for networks that can learn to discriminate human speech sounds. These networks perform equally well having been trained on speech spectra from any human language so far tested (English, Cantonese, Swahili, Farsi, Czech, Hindi, Hungarian, Korean, Polish, Russian, Slovak, Spanish, Ukranian and Urdu). Training the feature detectors requires exposure to just one minute of speech in any of these languages.

Description of the Model

The model builds on previous connectionist models, particularly the broad class of models known as interactive activation models, with three major modifications. Firstly, each network learns using many different, unsupervised learning rules. These use only local information, and so are biologically plausible. Secondly, every network is split into a number of separate subnetworks. This allows exploration of different neuronal architectures, and it becomes possible to use different learning rules to connect subnetworks. Each subnetwork has its own time-constant, and therefore responds to information in a specific range of time-scales. Finally, networks are evolved using a technique called genetic connectionism. Using a genetic algorithm allows great flexibility in the type of neural network that can be used. All the attributes of the neural network can be simultaneously optimised rather than just the connections. In this model the architecture, learning rules and time-constants are all optimised together.

The dynamics of all units in the network are first order and determined by summing activation from all connected units and making a change in activity proportional to the summed input activity, scaled by the time constant. Complex architectures can be represented in the model by using a subnetwork connectivity matrix that determines which learning rule will be used for the connections between any pair of subnetworks. If an element is zero there are no connections between two subnetworks. A positive integer element indicates that subnetworks are fully connected and the value of the integer specifies which one of the many learning rules to use for that set of connections as shown in figure 1.

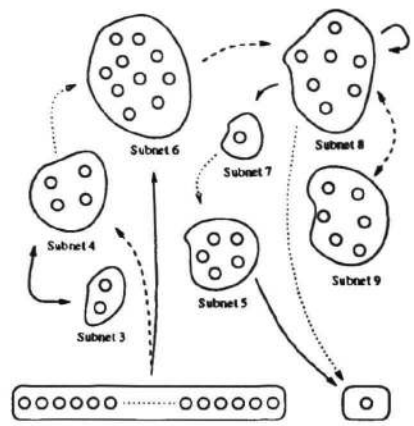

Figure 1: A network with 9 subnetworks. Subnetwork 1 and 2 are the input and output subnetworks, respectively. Arrows represent sets of connections and the type of learning rule employed by those sets of connections. There are three learning rules used; solid arrow (learning rule 1), dashed arrow (learning rule 2) and dotted arrow (learning rule 3).

Results

All of the human languages tested seemed to be equally effective for training the network to represent English speech sounds. To see whether *any* sounds could be used for training, the network was trained on white noise. This resulted in slower learning and a lower fitness. The fitness for a network trained on white noise never reached that of the same network trained on human speech.

The advantages of innately guided learning over conventional self-organising networks are that innate learning is much faster and is *less* dependent on the "correct" environmental statistics. It also offers an account of how infants from different linguistic environments can come up with the same featural representation so soon after birth. The model therefore demonstrates how genes and the environment could interact to ensure rapid development of a featural representation of speech.