SUMMARIES OF MASTER'S THESES

Classification of vowel segments using neural networks

Sander Kloosterman

Summary

The purpose of this study was to analyse the usefulness of neural networks in classifying vowel segments. We compared two different ways of representing a vowel, using either sixteen bandfilter values or three formant frequencies. An additional parameter was the fundamental frequency. We also examined the way information about a vowel is extracted from the parameters presented to the neural network.

The performance of a neural network depends both on its structure (topology) and the algorithm used in training the network. In this study we used the back-propagation algorithm. First a number of experiments have been carried out to learn more about the possibilities and impossibilities of different network topologies. A network of two layers appears to be sufficient for classification of vowel segments. It is possible to make a good estimation of the minimum number of elements required in each hidden layer.

Neural networks were trained using either formant frequencies or bandfilter values. We also tried to train some networks in which the fundamental frequency was used as an additional parameter, but this hardly improved the performance.

Using carefully chosen training-parameters, a model with four nodes in the hidden layer was able to classify 83.1% of al the presented vowels correctly. A model with a hidden layer consisting of six nodes scored 91.1%.

The vowel stimuli used in this experiment had also been used in a listening experiment, so the results of classification by a neural network can be compared to classification by human listeners. When comparing confusion-matrices of the listeners to those of neural networks, it appeared that the 'mistakes' made by the model were similar to human mistakes.

Kloosterman, S. (1992): 'Classification of vowel segments using neural networks'. Master's thesis, IFA report **119**, 71 pp. (in Dutch: 'Herkenning van klinkersegmenten met neurale netwerken').

Cellular Automata and Speech Recognition

Reinier Kortekaas

Summary

In this master's theses the applicability of cellular automata with respect to speech recognition was investigated. Von Neumann is generally considered to have initiated the study of cellular automata (Von Neumann, 1966). In a cellular automata system, the next state of a cell (at time t+1) results from both its own state and the states of cells in some neighbourhood (at time t). The number of distinct, possible states typically is strictly limited. The exploration of 1 dimensional, linear cellular automata by Wolfram has been the basis of our study. These discrete dynamical systems display complex behaviour despite their simple construction (e.g. fractal growth and, in Turing sense, universality; Wolfram, 1984). Cellular automata have been used to model a variety of phenomena e.g. in fluid dynamics and chemical reactions.

The main question in our study was whether cellular automata could produce 'characteristic patterns' when presented with 'characteristic inputs' (i.e. representations of static vowels). Starting point was to keep the cellular automata as simple as possible. A circular array of cells simulated a Bark scaled frequency axis. Cells had binary states indicating either activity or quiescence in a Bark scaled 'channel'. In this way vowels could be encoded by means of their formant frequencies (in Kortekaas (1992) only F1 and F2 were considered). A 'characteristic pattern' was generated by first initializing the array with such a vowel representation and then iteratively applying a particular transition function. Each newly evolved 'configuration' of cell states was added to the pattern. The same function operated on all cells, in a parallel fashion (The set of possible functions is completely determined by the cellular automaton specifications). In its state transition, each cell was only affected by its two adjacent 'neighbours'. In other words, only local characteristics of the array are 'processed' at each iteration.

For all average Dutch male vowels such a (binary) 'characteristic pattern' was generated and stored as a reference. Recognition of an unknown input vowel was executed by means of 'bitwise' comparison (under Hamming distance) of its evolved pattern and each of the stored references. For an array of 14 cells, patterns typically consisted of about 30 iterations. Often oscillatory behaviour was observed which determined the required pattern 'length'.

It was found that a particular state transition function favourably offered a fast means of 'characteristic pattern' generation ('totalistic rule 6' in Wolfram's terminology). This function displayed both spread of information along the array as well as spatio-temporal regularity. With this function it was possible to obtain a vowel triangle-like structure of recognitions (this result was reproduced for a number of lattice dimensions). However these vowel triangles displayed minor (local) irregularities. The intrinsic simplicity of the cellular automaton system made that these errors probably could not be easily removed (without actually changing the system itself). Moreover it remains unclear how such systems would behave when presented with more realistic speech input (i.e. dynamic input and additional noise disturbances). Future research could integrate some of these (essential) refinements. Besides, some alternative constructions of cellular automata systems seem to be forceful in speech recognition.

Kortekaas, R. (1992): 'Cellular Automata And Speech Recognition'. Master's thesis, IFA report **120**, 36pp.

Von Neumann, J. (1966): *Theory of Self-Reproducing Automata*, A.W. Burks (Ed.) University of Illinois Press, Urbana.

Wolfram, S. (1984): 'Universality and Complexity in Cellular Automata'. *Physica* **10D**, North-Holland, Amsterdam.