ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 83-88

Connectionist models investigating representations formed in the sequential generation of characters

F.M.Richardson [1], N. Davey [1], L.Peters [1], D.J.Done [2], S.H.Anthony [2]

F.1.Richardson, N.Davey, L.Peters, D.J.Done, S.H.1.Anthony @herts.ac.uk
Department of Computer Science [1], Department of Psychology [2],
University of Hertfordshire, College Lane, Hatfield, Hertfordshire, AL10 9AB

This paper considers the results of three different methods of encoding visual and motor representations of single sequential character production using three different architectures for the simulation of perceptual and motor processes. Examination of such processes through neural net modelling of the generation of handwritten characters promises to be a fruitful avenue of exploration as the induced representations of the models can be examined. The results of this analysis showed that both spatial and temporal similarity were important in these representations. Similar results have been shown to be true for actual representations in the motor cortex.

1. Introduction

Perception and action are closely connected. In recent psychological literature their relationship has developed to the point where they have become inseparable partners in cognitive function (Glenberg 1997). The mode of interaction between these two systems however remains unclear. Prinz (1997) discusses the 'common-coding approach' which advocates that perception and action share common codes and can communicate directly with each other, as opposed to the classical viewpoint that communication takes place via a translator. Hikosaka et al (1999) have proposed that two independent systems (which have parallel processes for the acquisition of spatial and motor co-ordinates) compete and interact during the learning of sequential movement. An example of such a sequential movement is handwriting.

In order to successfully generate a character both sequential (the order of character segments) and directional information (the direction of drawing character segments) is required. Such information requires representation at a cognitive level. From a motoric viewpoint a representation is required in the form of a movement plan (such as a generalised motor program) which encompasses information concerning the entire movement trajectory (including intermediate pen-up movements in the case of generating discontinuous character forms). From a visual perspective a representation is required in which sequential and directional information is encoded (such as the sequence of pen strokes), but without the explicit encoding of intermediate movements, (although it is possible for this information to be inferred).

To investigate the interaction of visual and motor representations in character generation it is necessary to define such representations within an encoding system that can account for the information held within the representations, and investigate the consequences of the use of different encoding approaches to modelling sequential generation. This paper investigates the internal representations formed as a result of using three different encoding strategies, which correspond to the perceptual and motor experience of character generation.

2. Models

2.1 Architectures

Representations were investigated using three Simple Recurrent Network (SRN) architectures, used because they have memory thus enabling generation of sequences. Networks used were the Elman, Jordan and a hybrid architecture, in which features of the first two were combined (and are displayed below in figure 1). All three architectures were investigated, as it was not clear which would be the most appropriate.

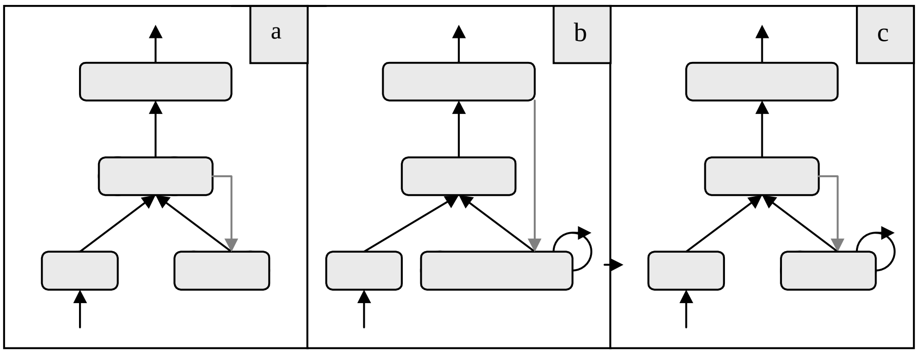

Figure 1: shows the (a) Elman network architecture (b) Jordan network architecture and (c) the hybrid network architecture used in the simulations. Black arrows indicate full connectivity between layers and grey arrows indicate one-to-one connections for the purpose of copying activation from the hidden/output layer to the context layer. Circular arrows indicate layer in which units have self-connections.

2.2 Output Encoding

All simulations were attempted using three forms of encoding. Output sequences forming letter shapes were generated as a result of specific single unit activation in the input layer. The size of the input layer corresponded to the number of sequences in the training set. Each letterform sequence and input unit was numbered. Activating the corresponding unit to the desired sequence generated an output sequence. This can be seen in figure 2. Output was represented on a 4 x 5-unit grid for networks with

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 83-88

a visual representation (see sections 2.2.1 and 2.2.2). Output for networks using a motor representation (see section 2.2.3) was in the form of two groups of 5 units, one group representing the x-axis and the other the y-axis of a co-ordinate system.

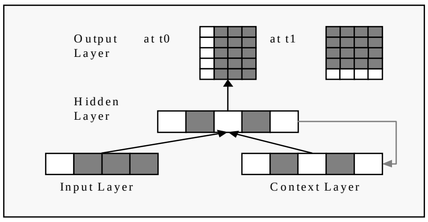

Figure 2: shows an example of the network architecture for an Elman net using a segment-by-segment encoding. Inactive units are shaded. It can be seen that activating a single unit in the input layer triggers a specific sequence.

2.2.1 Bit-by-bit (visual representation)

This encoding was inspired by Kosslyn et al (1988) and sequentially activates units in the output layer with the previous unit of activation (previous context) remaining active in the next step of the sequence. The resulting visual display emulates the gradual generation of an image (see figure 3) in which sequence and direction of production of the letter shape can be perceived and intermediate movements can be inferred.

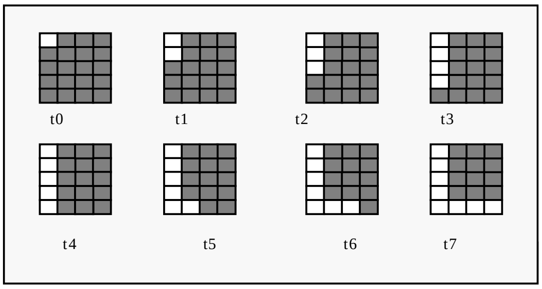

Figure 3: shows the output of a single sequence forming a letter L. Each square represents a unit on the 4 x 5 output grid. Inactive units are shaded. Activation at the current time step is added to that of the previous time step in order to generate a complete sequence.

2.2.2 Segment-by-segment (segmented visual target representation)

This encoding activates groups of units, with each group of units representing a letter segment but giving no directional information. This emulates the act of generating a character segment by segment. At each time step in the sequence a target segment is output the previous segment is not included in the current time step.

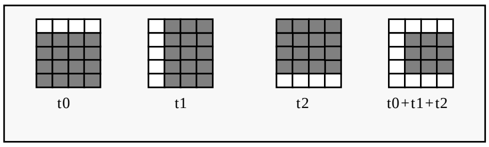

Figure 4: shows the output of a sequence forming a letter C. Each square represents a unit on the output grid. Inactive units are shaded. For this representation activation at the current time step does not include activation from the previous time step. A concatenation of each time step is shown, however this is not required output.

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 83-88

2.2.3 Thermometer encoding (movement trajectory representation)

The use of thermometer encoding was in order to encompass some representation of similarity into the co-ordinate system (Tijsseling and Harnad 1997). Unlike localist encoding, similar inputs have similar representations. In this case the movement from one target point to the next incorporates some notion of the direction of movement as well as the sequence. This form of encoding can account for movement trajectories (Georgopoulos 1997) as opposed to a visual display of the letterform. For example in the case of a discontinuous letterform such as X, the entire movement trajectory includes intermediate pen-up movements to another starting point.

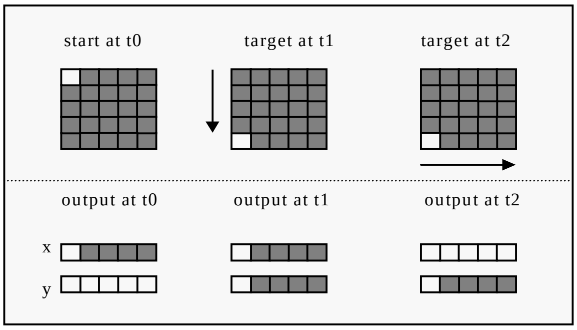

Figure 5: shows the x, y output sequence for a letter L. These co-ordinates are plotted on a grid for clarity. Inactive units are shaded. This encoding emulates the sequence of target points for generating a character. Arrows indicate movement towards the target point at the current time step from the previous time step.

3. Experiments

3.1 Training and testing

Training and testing patterns consisted of sequences that (according to the encoding method) generated a representation of letters C, L, G and J for the networks trained on bit-by-bit and segment-by-segment encoded patterns. For the thermometer networks training and testing patterns consisted of letters C, L, Z (continuous letter forms – no pen-up movement required) and E, F and X (discontinuous letter forms – intermediate pen-up movements required between segments). Batches of ten networks were run - each with randomly assigned weights. All networks were trained using the minimum number of hidden units which would allow consistent and successful learning of the task.

3.2 Results

All networks were trained successfully, after which they were able to produce complete sequences correctly. All models produced the same convergence behaviour and similar representations, when comparisons were made for each type of encoding. Principal component analysis (PCA) was conducted upon all units of the hidden layer every 10 epochs of training and upon completion of learning for a random selection of

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 83-88

three runs. An example of a PCA upon completion of training for each form of encoding is shown in figures 6, 7 and 8.

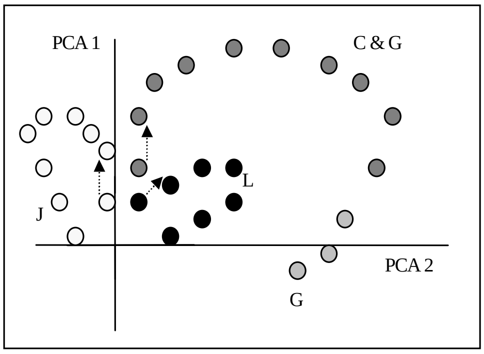

Figure 6: shows a 2-D PCA graph displaying the results of a PCA for a randomly selected Elman network upon completion of training, using bit-by bit encoding. It can be seen that sub-sequence C is represented within letter G. Sub-sequence L, however, which has a different starting point from letters G and C is not represented within letters C and G as a sub-sequence.

.

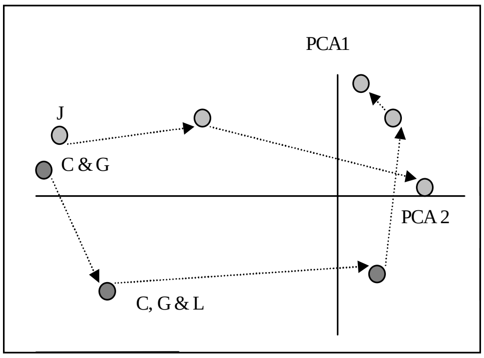

Figure 7: shows a 2-D PCA graph displaying the results of a PCA for a randomly selected Elman network upon completion of training, using segment-by segment encoding. It can be seen that sub-sequence C is represented within letter G. Also sub-sequence L is represented within letters C and G.

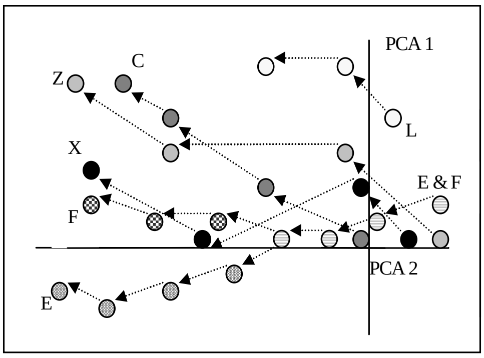

Figure 8: shows a 2-D PCA graph displaying the results of a PCA for a randomly selected Elman network upon completion of training using thermometer encoding. It can be seen that letters E and F, which share the same starting point share some representation, however letters C and L which do not share the same starting point do not.

In all cases spatial similarity was found to be internally represented. So for example letters C and G (see figures 6 and 7), where letter C is composed of the first three segments of letter G, are co-represented. Similarly letters F and E (see figure 8). A significant difference emerges between the bit-by-bit and thermometer encoding representations and the segment-by-segment encoding representations. The latter co-represents any segment independently of where it appears in the sequence of production. For example letters C and L (see figure 7), where the vertical segment of letter L appears as the second segment in the sequence forming the letter C, but are co-represented. However, the other two encodings ensure that the network only co-represents those sequences which are both temporally and spatially similar, so the C and L are not co-represented in this case.

4. Conclusions

Neural networks were used to investigate the implications of the task encoding upon the representations formed. The specific network architecture was not found to be of importance in terms of performance and representation. When input encoding provided spatial information only, then the induced internal representations were based on spatial similarity only. However, when input encoding included temporal information then the internal representation embodied temporal information too. These results reflect the strong trajectory specific relationship between spatial and temporal characteristics of letterforms (Viviani and Terzuolo 1982). Although providing a simplified model of the representation of sequential movement, these simulations nevertheless offer an explanation for how serial order and location of learnt sequences can be represented together in the motor cortex (Carpernter et al 1999). Further modelling work is required in order to determine which representation can best account for empirical data.

References

1. Carpenter, A.F., Georgopoulos, A.P. and Pellizzer, G. (1999) Motor cortical encoding of serial order in a context-recall task. *Science* **283,** 1752-1757.
2. Elman, J. (1990) Finding structure in time. *Cognitive Science* **14,** 179-211.
3. Georgopoulos, A.P. (1997) Neural networks and motor control. *The Neuroscientist* **3,** 52-60.
4. Glenberg, A. (1997) What memory is for. *Behavioural and Brain Sciences* **20,** 1-55.
5. Hikosaka, O., Nakahara, H., Rand, M., Sakai, K., Lu, X., Nakamura, K., Miyachi, S. and Doya, K. (1999) Parallel neural networks for learning sequential procedures. *TINS* **22,** 464-470.
6. Jordan, M.I. (1986) Attractor dynamics and parallelism in a connectionist sequential machine. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*. (Amherst 1986), 531-546. Hillsdale: Erlbaum.
7. Kosslyn, S., Backer Cave, C., Provost, D. and Von Gierke, S. (1988) Sequential processes in image generation. *Cognitive Psychology* **9**, 129-154.
8. Prinz (1997) Perception and action planning. *European Journal of Cognitive Psychology* **9,** 129-154.
9. Tijsseling, A., Harnad, S., (1997). Warping similarity space in category learning by backprop nets. Paper presented at the *Interdisciplinary Workshop on Similarity and Categorisation*, University of Edinburgh, November 1997.
10. Viviani, P. and Terzuolo., (1981). Trajectory determines movement dynamics. *Neuroscience* **7,** 431-437.