ESANN'2007 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2007, d-side publi., ISBN 2-930307-07-2.

Sparsely-connected associative memory models with displaced connectivity

Lee Calcraft, Rod Adams and Neil Davey

School of Computer Science, University of Hertfordshire
College Lane, Hatfield, Hertfordshire, AL10 9AB, U.K.
{L. Calcraft, R.G. Adams, N. Davey} @ herts.ac.uk

Abstract. Our work is concerned with finding optimum connection strategies in high-performance associative memory models. Taking inspiration from axonal branching in biological neurons, we impose a displacement of the point of efferent arborisation, so that the output from each node travels a certain distance before branching to connect to other units. This technique is applied to networks constructed with a connectivity profile based on Gaussian distributions, and the results compared to those obtained with a network containing purely local connections, displaced in the same manner. It is found that displacement of the point of arborisation has a very beneficial effect on the performance of both network types, with the displaced locally-connected network performing the best.

1 Introduction

In recent work [1, 2] we have explored the pattern-completion performance of sparsely-connected associative memory models using a variety of different connection strategies. Our studies included networks in which the probability of a connection between any two nodes varied with distance according to Gaussian and exponential probability distributions. We concluded that when wiring costs are taken into account, relatively tight Gaussian and exponential connectivity distributions perform the best, and are considerably more efficient than small-world connection strategies based on the progressive rewiring of locally-connected networks.

In the present work, we explore the performance of sparsely-connected associative memories using a new connection strategy: that of displaced connectivity. This work is inspired by the work of Hertzog et al [3] on the synchronous spiking behaviour of neurons, in which they explore the use of local connectivity with a small lateral displacement, in a 2D neural 'gas'.

Their work was concerned with establishing, in a locally-connected network, the displacement distance which would provide a compromise between mean connection length and the mean minimum path length (the minimum number of steps separating each pair of nodes averaged over the whole network), as used by Watts and Strogatz [4]. Our study is concerned with the effect of lateral displacement on the pattern-completion performance of 1D associative memory models with local connectivity, and also with patterns of connectivity based on a Gaussian distribution.

2 Network dynamics, training and performance measurement

A network of perceptrons is arranged in a one-dimensional structure with wrap-around at the ends (see Figure 1a), and is trained on sets of random patterns of length

N, where N is the number of units in the network. The output of each unit is connected to the inputs of a fixed number, k, of other units. The networks used in the present studies have no symmetric connection requirement [5], and the recall process uses asynchronous random order updates, in which the local field of unit i is given by:

$$h_i = \sum_{j \neq i} w_{ij} S_j$$

where w_{ij} is the weight on the connection from unit j to unit i, and S $(= \pm 1)$ is the current state. The dynamics of the network is given by the standard update: $S_i' = \Theta(h_i)$, where Θ is the Heaviside function. Network training is based on the perceptron training rule [6] chosen for its higher resultant capacity than that of the standard Hopfield model. The rule is designed to drive the local fields of each unit the correct side of the learning threshold, T, for all the training patterns. Earlier work has established that a learning threshold of $T = 10$ gives good results [7].

Network performance is determined by measuring Effective Capacity [8] [9]. This is a measure of the number of patterns which a network can restore under a specific set of conditions. The network is first trained on a set of random patterns. Once training is complete, the patterns are each randomly degraded with 60% noise, before presenting them to the network. After convergence, a calculation is made of the degree of overlap between the output of the network, and the original learned pattern. The Effective Capacity of the network is the highest pattern loading at which this mean overlap for the pattern set is 95% or greater. The Effective Capacity of a network has been shown to track its underlying maximum theoretical capacity [9].

3 Gaussian connectivity with displaced efferent arborisation

In the first part of our study we examine the performance of a network with a Gaussian connectivity distribution in which the point of efferent arborisation is given a progressively larger lateral offset. The biological basis for this model would be an axon emerging from each neuron, which then travels for a distance (equivalent to our

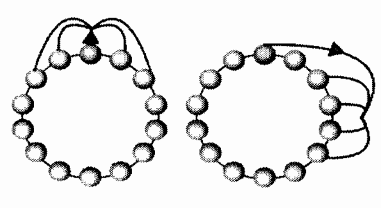

Fig. 1: Illustration of a locally-connected network, a, and a network with displaced local connectivity, b. Both networks have 14 nodes, with 4 efferent connections per node.

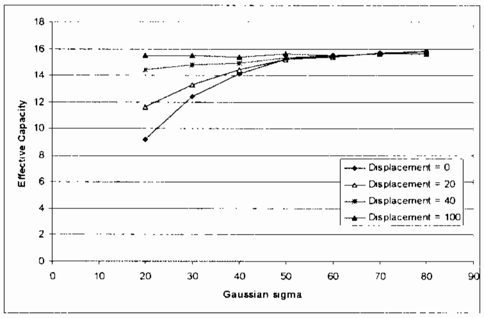

Fig. 2: Effective Capacity vs Gaussian σ in networks with a range of arborisation displacements, from 0 to 100. The network is built from 500 units with 50 efferent connections per node. Results are averaged over 200 runs.

displacement distance) before arborisation. It then connects to other neurons with a Gaussian probability distribution of width σ. See Figure 1b, which illustrates the displacement of a small locally-connected network. In our simulations we give a random direction of travel to the emergent axon around the ring, so that it has an equal probability of travelling clockwise or anticlockwise around the ring before arborisation.

Figure 2 shows the results obtained with a network of 500 units with 50 efferent connections per node. Measurements of Effective Capacity were made for the following lateral displacements: 0, 20, 40 and 100 units, varying the Gaussian σ within the range 20 to 80 in each case. As may be seen, the effect of introducing a progressively larger lateral displacement is to decrease the spread of the Gaussian required to produce a particular Effective Capacity. So, for example, to achieve an Effective Capacity of around 12, the non-displaced Gaussian requires a σ approaching 30, while the Gaussian displaced by 20 units only requires a σ of around 20 to achieve a similar result. In other words, the greater the displacement, the less broad the required spread of connectivity at the point of arborisation to achieve a given performance. When the displacement reaches about 100 units, there appears to be no further advantage in broadening the Gaussian spread of connections (the maximum possible displacement around the ring is 250 units).

Although we can see from Figure 2 that all four plots achieve the same maximum Effective Capacity of around 16, and that those with the greatest displacement reach the maximum value at lower values of σ, the effect of displacement on wiring cost is not immediately obvious. To clarify this, Figure 3 shows the Effective Capacity of the four networks plotted against the corresponding mean wiring length of each network. In calculating wiring length here, we assume that the output of each node travels along a single connection until it arborises.

For comparison purposes, Figure 3 also depicts the results for a non-displaced progressively-rewired network (rewired in steps of 10% from 0 to 100%), which, as we have previously shown [1], performs less well than its Gaussian counterpart, and is the poorest performer here.

From the graph it may be seen that while all architectures are capable of attaining the highest Effective Capacity of about 16, there are considerable differences in wiring costs. The progressively-rewired network performs the worst, while the Gaussian networks show better and better performance as the arborisation point is progressively displaced. The network whose points of arborisation are displaced by 100 units achieves an Effective Capacity of around 15.5 at a mean wiring length of just 19 units, for example, whereas the non-displaced Gaussian reaches a mean wiring length of 60 before it achieves the same Effective Capacity. This represents a considerable gain.

4 Local connectivity with displaced efferent arborisation

The results in Figure 3, in which very tight Gaussian distributions with displaced arborisation perform well, suggest that we may also be able to obtain good performance by displacing the arborisation point in a network with purely local

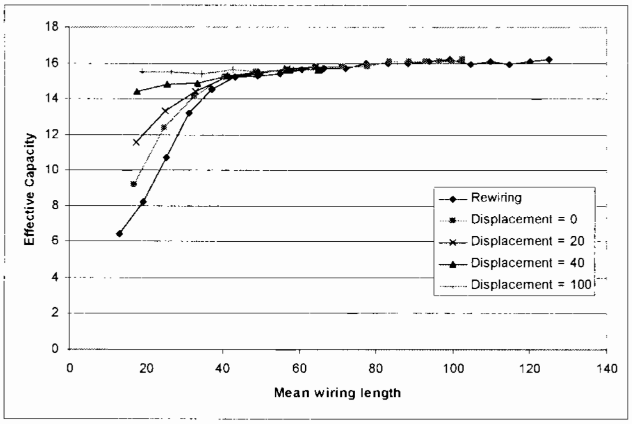

Fig. 3: Effective Capacity vs mean wiring length for a network of 500 units with 50 efferent connections per node. Architectures are based on a progressively-rewired network (rewired in steps of 10% from 0 to 100%), together with four networks with Gaussian distributions of varying width, each of whose point of efferent arborisation has a different lateral displacement, *d*, ranging from 0 to 100 units. Results are averaged over 200 runs.

connectivity. To examine this question we measured the Effective Capacity of a locally-connected network of 500 units, each with 50 efferent connections, with lateral displacements in the range 0 to 200 units. Figure 4 shows the results. The first point on the graph represents the performance of the network with purely local connectivity (displacement is zero). This is poor, as expected. The first displacement of 10 units brings a small improvement, but then successive increases reap more considerable rewards, with a reasonably steep linear increase in Effective Capacity as the displacement is increased from 10 to 50 units in steps of 10. The response then flattens out at an Effective Capacity of around 16 by the time the displacement has reached 60 or 70.

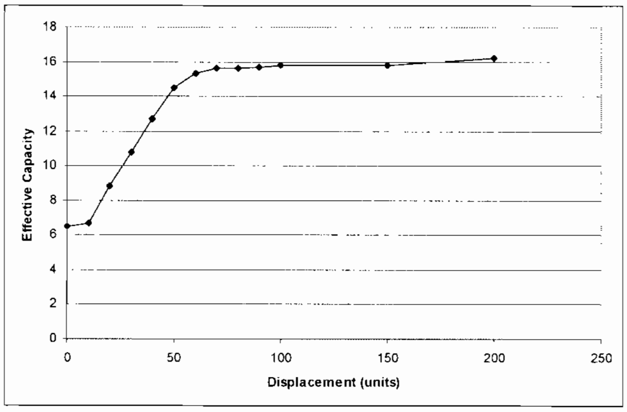

Fig. 4: Effective Capacity vs displacement for a locally-connected network of 500 units, each with 50 afferent connections. Results are averages over 200 runs.

In order to see the effect of displacement of a locally-connected network on performance when wiring costs are taken into account, we again plot Effective

ESANN'2007 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2007, d-side publi., ISBN 2-930307-07-2.

Capacity against mean wiring length; and for comparison purposes we have included the displaced Gaussian and rewiring data from Figure 3. The results are shown in Figure 5.

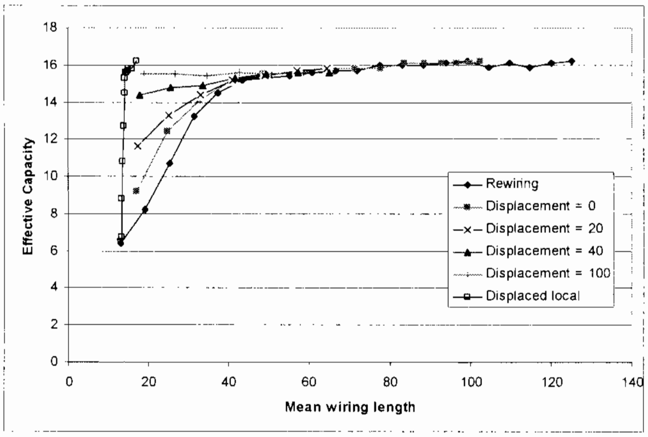

Fig. 5: Effective Capacity vs mean wiring length for a network of 500 units, with 50 efferent connections per node. Architectures are based on a local network with increasing displacement from 0 to 100, a progressively-rewired network, and four networks with Gaussian distributions of varying width, each of whose point of efferent arborisation has a different lateral displacement, *d*, ranging from 0 to 100 units. Results are averaged over 200 runs.

As may be seen, the displaced-local network clearly outperforms all the others, reaching an Effective Capacity of more than 16 at a mean wiring length of 17 units. None of the other networks reach this level of Effective Capacity until their mean wiring lengths are considerably greater.

The displaced-local network thus appears to be an interesting contender when attempting to achieve good pattern-completion performance at low wiring costs. Moreover, the time which the associator takes to converge during recall appears to be comparable to other well-performing networks of its size (about 8 epochs in the above tests).

The reason for the improved wiring efficiency observed here in networks built with displaced connectivity appears to lie in the sharing of the efferent conduit. In order for an associative memory to perform well, each of its nodes must be connected to some of the nodes which are not immediately local to it [2]. This requirement for non-local connectivity, however, significantly adds to the mean wiring length of the network. But by using displaced efferent arborisation, non-local connectivity can be achieved at a very low wiring cost.

5 Conclusion

Our experiments establish the important result that by displacing the point of efferent arborisation in a sparsely-connected associative memory model, we can significantly reduce wiring costs while achieving the same pattern completion performance.

In networks with patterns of connectivity based on a Gaussian distribution, the effect of introducing a displacement of the point of efferent arborisation is to increase the Effective Capacity of the network. The improvement in performance is most noticeable with tight Gaussian distributions ($\sigma = 20$), where introducing a displacement of 20 units results in an increase in Effective Capacity from around 9 to just below 12. The overall effect of introducing a displacement of efferent arborisation in a network with Gaussian connectivity is to decrease the mean wiring length of the network at which a particular Effective Capacity is reached. In this respect the best results are achieved with the tightest Gaussian distributions and the largest displacements.

In the second set of experiments, the performance of a locally-connected network was measured at progressively greater displacements of the point of efferent arborisation. It was found that as the displacement was increased from 10 to 50 units, the Effective Capacity increased linearly from around 6 to 15, after which little further improvement occurred. In terms of achieving a high Effective Capacity at low wiring costs, the displaced local network performed the best, exceeding the results of the displaced Gaussian. At a displacement of 70 units the Effective Capacity of the displaced local network is 15.6, and has a mean wiring length of 14.4 units. Only the very tightest Gaussian distribution ($\sigma = 20$) with the greatest displacement (displacement = 100 units) came close to this, with an Effective Capacity of 15.5 and a mean wiring length of 19.0 units.

It thus appears that a network built using purely local connectivity with displaced efferent arborisation represents an exceedingly efficient connection strategy for the construction of associative memories. This strategy would be expected to pay dividends in any physical implementation of associative memories such as those considered here, whether biologically or silicon based.

References

[1] L. Calcraft, R. Adams, and N. Davey, "Gaussian and exponential architectures in small-world associative memories," *Proceedings of ESANN 2006: 14th European Symposium on Artificial Neural Networks. Advances in Computational Intelligence and Learning*, pp. 617-622, 2006.

[2] L. Calcraft, R. Adams, and N. Davey, "High Performance Associative Memory Models with Low Wiring Costs," *Proceedings of 3rd IEEE Conference on Intelligent Systems, University of Westminster, 4-6 September 2006-09-29*, pp. 612-616, 2006.

[3] A. Herzog, K. Kube, B. Michaelis, A. D. de Lima, and T. Voigt, "Connection strategies in neocortical networks," *Proceedings of ESANN 2006: 14th European Symposium on Artificial Neural Networks. Advances in Computational Intelligence and Learning*, pp. 215-220, 2006.

[4] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440-442, 1998.

[5] N. Davey, L. Calcraft, and R. Adams, "High capacity, small-world associative memory models," *Connection Science*, vol. 18, pp. 247-264, 2006.

[6] S. Diederich and M. Opper, "Learning of correlated patterns in spin-glass networks by local learning rules," *Physical Review Letters*, vol. 58, pp. 949-952, 1987.

[7] N. Davey, S. P. Hunt, and R. G. Adams, "High capacity recurrent associative memories," *Neurocomputing*, vol. 62, pp. 459-491, 2004.

[8] L. Calcraft, "Measuring the performance of associative memories," University of Hertfordshire Technical Report (420) May 2005.

[9] L. Calcraft, R. Adams, and N. Davey, "Locally-connected and small-world associative memories in large networks," *Neural Information Processing - Letters and Reviews*, vol. 10, pp. 19-26, 2006.