ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

Gaussian and Exponential Architectures in Small-World Associative Memories

Lee Calcraft, Rod Adams and Neil Davey

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

Abstract. The performance of sparsely-connected associative memory models built from a set of perceptrons is investigated using different patterns of connectivity. Architectures based on Gaussian and exponential distributions are compared to networks created by progressively rewiring a locally-connected network. It is found that while all three architectures are capable of good pattern-completion performance, the Gaussian and exponential architectures require a significantly lower mean wiring length to achieve the same results. In the case of networks of low connectivity, relatively tight Gaussian and exponential distributions achieve the best overall performance.

1 Introduction

It has been shown that the small-world class of networks in which the majority of nodes are connected to their nearest neighbors, but in which a proportion of connections are distal, display interesting properties [1]. In such networks, the degree of clustering remains almost as high as that of a locally-connected network, while the distal connections of each node are sufficient to maintain a relatively low mean minimum path length across the network.

Bohland and Menai [2] and Davey et al. [3] have both applied these principles to the design of small-world associative memory models. In each case a one-dimensional lattice in the form of a ring was established as a locally-connected sparse associative memory, and its performance was measured, as the local connections were progressively rewired to randomly selected nodes. It was found that the performance of the network steadily increased with rewiring, up to the point where 40-50% of connections had been rewired. Beyond this point, further rewiring had little additional effect on performance.

In the present paper, we explore the effect of using alternative connectivity architectures based on Gaussian and exponential distributions, and compare the resulting pattern-completion performance with that of an initially locally-connected network which is progressively rewired as above.

2 Network dynamics and training

A network of perceptrons is arranged in a one-dimensional structure with wrap-around at the ends, and is trained on sets of random patterns of length N, where N is the number of units in the network. The input of each unit is connected to the outputs of a fixed number, k, of other units. The networks used in the present studies have no symmetric connection requirement [4], 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 [5] 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 [6].

3 Performance measurement

Network performance is determined by measuring the Effective Capacity [7] [8], and all results have been verified against the mean radius of the basins of attraction [9] as an alternative performance indicator. Effective Capacity 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. This is repeated for each pattern in the set, and a mean overlap for the whole pattern set is calculated. The Effective Capacity of the network is the highest pattern loading at which this mean overlap is 95% or greater. This measure affords certain advantages over the radius of the basins of attraction, not the least of which is its lack of upper bound, and the way in which it tracks the underlying maximum theoretical capacity of a network [7] [10].

In physical systems, whether biological or electronic, the length of wiring between nodes will be an important issue, and we take account of this by plotting the network's pattern restoring ability, as measured by Effective Capacity, against mean wiring length, for each variant of each generic network type under test.

4 Gaussian, exponential and rewiring architectures compared

A network of 5000 units, each with 50 afferent local connections was set up as a one-dimensional lattice. It was trained on sets of random patterns using perceptron training rules, and its Effective Capacity measured. 10% of the network connections were then rewired to random connection points around the ring, and the network retrained, and retested. This procedure was repeated for different levels of rewiring in steps of 10% up to 100%, this latter representing a fully-random network. The network was then rebuilt with a Gaussian connectivity distribution, so that the probability of a connection between any two nodes separated by a distance d $(1 \leq d < N/2)$ around the ring was proportional to

ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

$$\frac{1}{\sigma}\exp\left(-\frac{(d-1)^2}{2\sigma^2}\right)$$

and measurements of Effective Capacity were made for a series of such networks, whose σ ranged in value from 10 to 2500. This process was then repeated, using progressively tighter exponential distributions, where the probability of connection was proportional to $\exp(-\lambda(d-1))$, with λ in the range 0.001 to 0.06.

Figure 1 shows the resultant Effective Capacity of all the iterations of the three types of network, plotted against the mean wiring length of each network configuration. It is apparent from this that while all three generic architectures are capable of achieving the highest Effective capacity of 23, the Gaussian and the exponential distributions do so at a far lower wiring cost. In order to achieve an Effective Capacity of 22, the rewired network would need to be 50% rewired, and this results in a mean wiring length of 630, while the Gaussian (at a σ of 120) and the exponential network (at a λ of 0.01) would both have a mean wiring length of just 96, and are thus very considerably more efficient in terms of achieving high Effective Capacity at low mean wiring length. The extreme closeness of the Gaussian and exponential plots in Figure 1 is also worthy of note, and is discussed later.

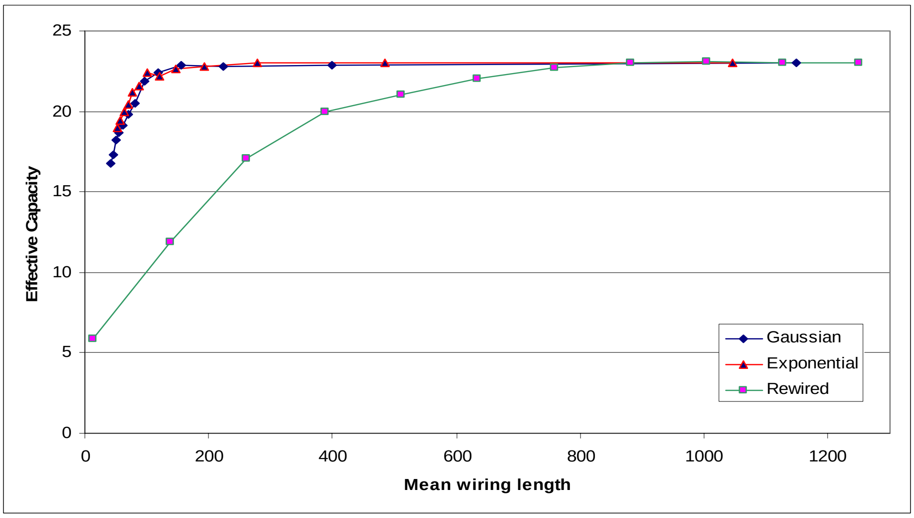

Fig. 1: Effective Capacity against wiring length for a network of 5000 units with 50 afferent connections per unit (1% connectivity). Comparison of Gaussian, exponential and rewiring architectures. Results are averages over 5 runs for each connectivity distribution.

To shed light on these results, Figure 2 shows the connectivity profiles of the three network configurations which achieve the same Effective Capacity of 22. From this we can see how tightly clustered are the connectivity distributions for the successful Gaussian and exponential variants. Even with their asymptotic tails, these distributions will have few, if any, long distance connections, and yet they still achieve very good pattern completion performance. The rewired network clearly has a relatively large number of long distance connections. These are, perhaps surprisingly, not essential to performance, and yet it is these which are responsible for the high mean wiring length of the network.

4.1 Comparing networks with higher connectivity levels

In networks with higher levels of connectivity, the difference in performance between the three generic architectures is much less marked. Figure 3 shows a similar plot of Effective Capacity against wiring length to that of Figure 1, but for a network of 500 units with 50 connections per unit: ten times the connectivity of the previous networks. The Gaussian and exponential results are again inseparable. But at this higher level of connectivity, the performance of the rewired network is much closer to that of its Gaussian and exponential counterparts, though both still outperform the rewired network in the important region of wiring lengths between 20 and 60, where the Effective Capacity is relatively high, and the mean wiring length relatively low.

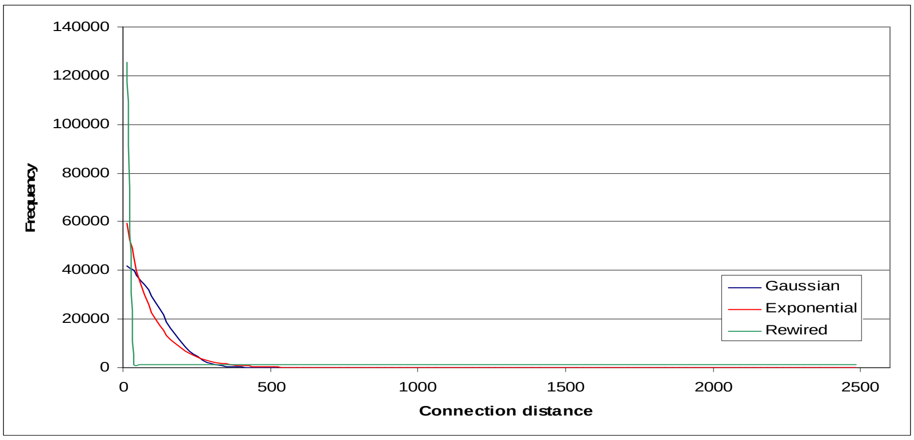

Fig. 2: Connectivity histogram for a network of 5000 units, each with 50 connections (1% connectivity), comparing Gaussian, exponential and rewiring architectures with the same Effective Capacity of 22.

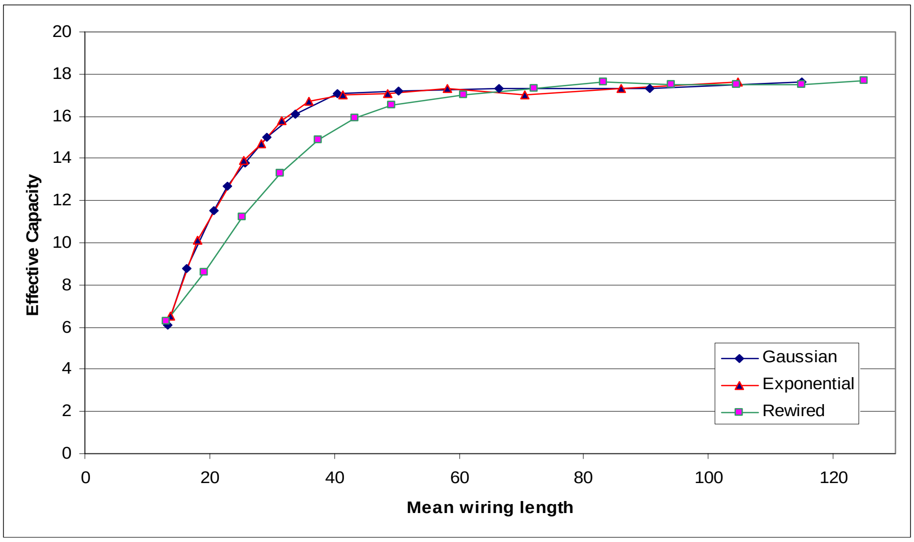

Fig. 3: Effective Capacity against wiring length for a network size of 500 units with 50 afferent connections per unit (10% connectivity). Comparison of Gaussian, exponential and rewiring architectures. Results are averages over 50 runs for each connectivity distribution.

If we take points of similar Effective Capacity from this graph, we can again compare specific connectivity distributions. We will use the points corresponding to an Effective Capacity of 16.1 on the Gaussian network (at a σ of 42), 15.8 on the exponential (at a λ of 0.032), and 15.9 on the progressively-rewired network (at a rewiring of 25%). These have mean wiring lengths of 33.8, 31.6 and 43.3 respectively. The connectivity distributions for these architectures are shown in Figure 4, and although the three distributions differ significantly, the connectivity profile of the progressively-rewired network is closer to the Gaussian and exponential profiles than it was in the 5000 unit network with only 1% connectivity shown in Figure 2. This rapprochement of the rewired connectivity profile to that of the Gaussian and exponential is reflected in the increased efficiency of the rewired network, though as already mentioned, it is still less efficient than its Gaussian and exponential counterparts in terms of Effective Capacity at low mean wiring length.

It is also of interest here that in spite of the clear differences between the Gaussian and exponential profiles, their pattern-completion performance is indistinguishable from each other. One might have expected that the Gaussian's relatively smaller number of very close connections, or its slightly larger number of intermediate range connections might have given rise to a performance differential.

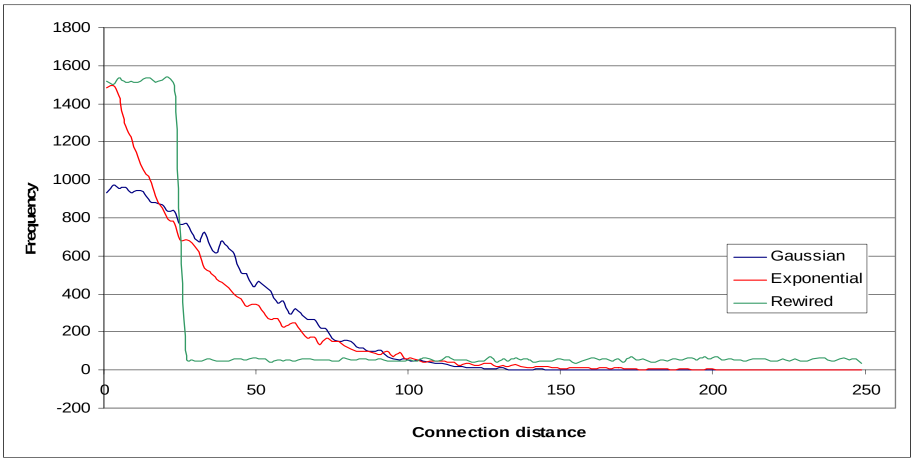

Fig. 4: Connectivity histogram for a network of 500 units, each with 50 connections (10% connectivity), comparing Gaussian, exponential and rewired architectures giving similar values of Effective Capacity (16.1, 15.8 and 15.9 respectively).

5 Conclusion

In sparsely-connected associative memories, patterns of connectivity based on Gaussian and exponential probability distributions are capable of a pattern-completion performance which far exceeds that of locally-connected networks, and which can match that of the best-performing random network. But the Gaussian and exponential architectures achieve this at a much lower mean wiring length than that of a randomly-rewired small-world network, or of a random network. This performance differential is greatest in networks with very sparse connectivity, as evidenced in the

1% connectivity network of 5000 units with 50 afferent connections per node. But even at connectivity levels of 10% (500 units with 50 afferent connections per node), the differential is still significant.

There was, however, no perceptible performance difference between the Gaussian and exponential networks right across the range of distributions, and at both the 1% and 10% connectivity levels tested. This is somewhat surprising in view of the difference in profile of the two distributions. One might have thought that the lower number of local connections, and the higher number of medium-length connections of the Gaussian relative to its exponential counterpart, might have resulted in a performance differential.

Turning now to the specific distributions which resulted in the best performance, it was found that in the sparsest networks, relatively tight Gaussian and exponential distributions achieved very good pattern-completion performance. A Gaussian distribution with a σ of just 120, and an exponential with a λ of 0.011 both achieved an Effective Capacity of 22. When one considers that the maximum connection length of the network is 2500 connections either side of any given node around the ring, it is clear that these are very tight distributions indeed. Thus, while connectivity must not be purely local, each node requires almost no distal connections, with the vast majority of connections being clustered around the host node in a relatively tight Gaussian or exponential distribution.

When connectivity is less sparse, it was found that the Gaussian and exponential architectures still perform the best, but that the optimal distributions are now somewhat less tight in proportion to the overall size of the network. We are currently studying the factors which determine the optimal distributions for differing network sizes and connectivity ratios.

References

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

[2] J. Bohland and A. Minai, "Efficient associative memory using small-world architecture," *Neurocomputing*, vol. 38-40, pp. 489-496, 2001.

[3] N. Davey, B. Christianson, and R. Adams, "High capacity associative memories and small world networks," presented at IJCNN, Budapest, 2004.

[4] N. Davey, L. Calcraft, and R. Adams, "High capacity small-world associative memory models," *To be published in Connection Science*.

[5] 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.

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

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

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

[9] I. Kanter and H. Sompolinsky, "Associative recall of memory without errors," *Physical Review A*, vol. 35, pp. 380-392, 1987.

[10] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*. Redwood City, CA: Addison-Wesley Publishing Company, 1991.