High Capacity Associative Memory Models with Bipolar and Binary, Biased Patterns

Weiliang Chen
University of Hertfordshire
w.3.chen@herts.ac.uk

Rod Adams
University of Hertfordshire
r.g.adams@herts.ac.uk

Lee Calcraft
University of Hertfordshire
l.calcraft@herts.ac.uk

Volker Steuber
University of Hertfordshire
V.Steuber@herts.ac.uk

Neil Davey
University of Hertfordshire
n.davey@herts.ac.uk

Abstract

The high capacity associative memory model is interesting due to its significantly higher capacity when compared with the standard Hopfield model. These networks can use either bipolar or binary patterns, which may also be biased. This paper investigates the performance of a high capacity associative memory model trained with biased patterns, using either bipolar or binary representations. Our results indicate that the binary network performs less well under low bias, but better in other situations, compared with the bipolar network.

1 Introduction

The functionality of associative memory which emerges in the mammalian cortex can be simulated using a single layer, recurrent neural network (Hopfield, 1982). In these models a training set of patterns is leant, so that the trained network will have these patterns as some of the fixed points of its dynamics. The *capacity* of the network is the maximum number of random patterns that it can learn as fixed points.

The canonical version of these models, named the Hopfield net, which uses a bipolar pattern representation (+1/-1) and one shot Hebbian learning, is known to have a low capacity and particularly poor performance when the training patterns are correlated. Given a network with N units, the theoretical maximum capacity of the canonical Hopfield model is approximately $0.14N$ (for non-correlated patterns). Another critical drawback of this type of associative memory model is that there is no guarantee that the training patterns are memorized.

Gardner (1988) introduced another associative memory model which used a perceptron type learning algorithm. This model provides a significantly higher maximum capacity, which is up to $2N$ for uncorrelated patterns, and actually increases with bias in the training set (Gardner, 1988), whilst still guaranteeing that all training patterns are memorized.

The investigation of high capacity associative memory models trained with biased patterns (patterns in which the probability of +1 occurring is not 0.5), using either bipolar or binary (1/0) representations is interesting for three reasons. Firstly, when compared with the bipolar representation, the binary representation is more biologically plausible as it does not assume negative neural activity. Secondly, activity in the mammalian brain is known to be sparsely coded (Braitenberg & Schüz, 1998). Finally, although the theoretical capacity of the network with biased, bipolar patterns is already known (Gardner, 1988), the capacity and performance of networks trained with binary, biased patterns are still unknown. It is surprising that no one, up to now, has investigated this topic experimentally. This paper gives the first experimental results on this topic. Results indicate that the binary network performs less well when the training set have low bias, but better in other situations, when compared with the bipolar network.

2 Details of Model Investigated

2.1 Bipolar and Binary Representations

The Hopfield model usually uses a bipolar representation. However it is also possible to construct a binary network. In the standard Hopfield model, these two representations can be shown to be functionally equivalent (Amit, 1989), though the choice of representation can affect the speed and efficacy of the learning algorithm. For example, using a binary representation together with Hebbian learning, the network can have only half of the capacity of the same size of network with a bipolar representation (Hopfield, 1982).

The simple perceptron learning rule is quite different when the patterns to be learnt are binary as opposed to bipolar. With binary patterns, learning only takes place on active connections, that is on afferent connections from units in the +1 state. In the bipolar case learning takes place on all incoming connections. However, a previous study (Davey et al. 2004) showed that there is no significant difference between networks with these two representations in performance when trained with unbiased patterns, although the binary network takes significantly longer to train.

The situation may be different when combining biased patterns with the bipolar or binary representations. The capacity of bipolar network with highly bias training patterns is known to tend towards infinity (Gardner, 1988). However, the capacity of an analogous using a binary representation is still in question.

2.2 Bias of the Patterns

Investigations into associative memory models usually assume unbiased training patterns. Formally the bias of a training set is the probability that any given bit is +1. That is, $prob(\xi = +1) = bias$, given ξ as the state of a unit in a training pattern. The restriction of unbiased patterns is useful for mathematical simplicity, but is often neither biological plausible nor practical. First of all, evidence from neuroscience (Abeles, 1982; Braitenberg & Schüz, 1998) indicates that the mean firing rate of the cortex is significant less that 50%, suggesting pattern activity with low bias. Secondly, in empirical areas such as image recognition, the patterns tend to be biased.

In the experiments reported here the training patterns are given a bias ranging from 0.1 to 0.9.

2.3 High Capacity Associative Memory Model

A description of the high capacity associative memory model is now given. The model uses two processes: training and network dynamics.

To train a network of perceptrons to act as an associative memory, the input and output layers consist of the same set of neurons. The weights can then be trained using any perceptron training procedure, so that the network autoassociates. See Figure 1.

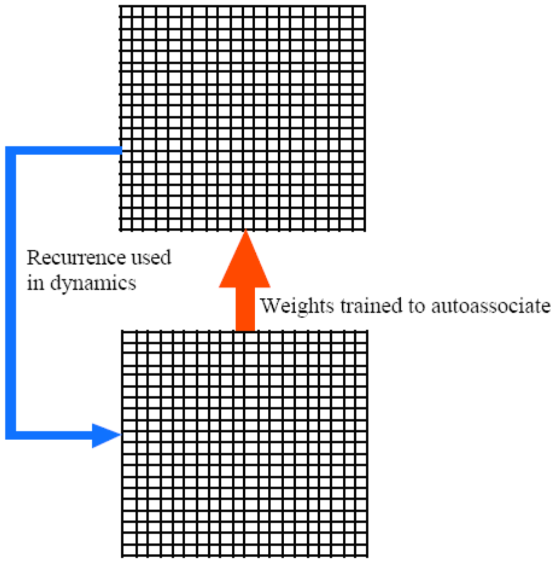

Figure 1. An abstract model of perceptron training. The red arrow represents the weights in an autoassociator of perceptrons. The blue arrow represents the recurrence of dynamics. The network will change states until a fixed point is reached.

Given a network of N units and a set of N-ary, training patterns, $\{\xi^p\}$, $\xi^p = [\xi_0^p, \xi_1^p, \xi_2^p, ..., \xi_N^p]$, the model learns patterns by modifying the N by N weight matrix denoted by **W**. After training, a specific pattern of unit states is first presented to the network. The network state is then modified according to an update rule that defines the network dynamics, until ending up with a stable state.

Denoting the weight of the connection from unit j to unit i in W by w_{ij}, the training of the Gardner model modifies all w_{ij} iteratively based on the unit state ξ_i^p and the unit's net input h_i^p, the *local field*, given by $h_i^p = \sum_{j \neq i} w_{ij} \xi_j^p$, together with a non-negative parameter, the *learning threshold*, denoted by T. The whole process of training can be described as:

```
Begin with a zero weight matrix
Repeat until all units are correct
 Set the state of the network to one of
the ξ^p
  For each unit, i, in turn:
   Calculate its local field h_i^p
   If ( ξ_i^p = on and h_i^p < T )
        or ( ξ_i^p = off and h_i^p > -T )
   then change the weight on connections
   into unit i according to:
```

$$\forall i \neq j \quad w'_{ij} = w_{ij} + \frac{\xi_j^p}{N},$$

When (ξ_i^p = *on* and $h_i^p < T$)

$$\forall j \neq i, \quad w'_{ij} = w_{ij} - \frac{\xi_j^p}{N},$$

When (ξ_i^p = *off* and $h_i^p > T$)

ξ_i^p = *on* denotes the *i*th bit of pattern *p* being +1
ξ_i^p = *off* denotes the value is -1 (bipolar) or 0 (binary)

Note that there are significant differences in training between a bipolar representation and binary representation. In the formula

$$w'_{ij} = w_{ij} + \frac{\xi_j^p}{N} \text{ and } w'_{ij} = w_{ij} - \frac{\xi_j^p}{N}$$

ξ_j^p is `off` means $\xi_j^p = -1$ in a bipolar network, but $\xi_j^p = 0$ in a binary network. This indicates that the training of a binary network only takes place on the afferent (incoming) connections from the units with +1 state, whilst the training of a bipolar network takes place on all afferent connections. Therefore the training of a binary network is expected to be a lot longer than the one of a bipolar network.

In the dynamics of this model, the changes of unit states are given by:

$$S'_i = \begin{cases} on & \text{if } h_i > 0 \\ off & \text{if } h_i < 0 \\ S_i & \text{if } h_i = 0 \end{cases} \qquad \text{where } S'_i \text{ is the new state of } S_i$$

.

The update of unit states can be either synchronous or asynchronous. In our experiments we use asynchronous random update. In the traditional Hopfield network, the asynchronous update as well as the symmetric weight matrix guarantee that the network state can be released to a fix point (Hopfield, 1982). However, the model in our experiments has no symmetric weight matrix. Nevertheless, the network almost always converges to a fixed point. If a pattern is in one of the fixed points of the network then this pattern is successfully stored and is considered a fundamental memory.

3 Experiments and Results

3.1 The Measure of Effective Capacity

To measure performance, we are interested in not only the actual capacity of the network, but also the network's ability to correct noisy patterns. Therefore the Effective Capacity (*EC*) (Calcraft, 2005; Calcraft, 2006) of the network is used in this paper. 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.

The Effective Capacity of a particular network is determined as follows:

```
Initialise the number of patterns, P,
to 0
Repeat
Increment P
  Create a training set of P random
  patterns
  Train the network
  For each pattern in the training set
      Degrade the pattern randomly by
      adding 60% of noise
      With this noisy pattern as start
      state, allow the network to
      converge
      Calculate the overlap of the final
      network state with the original
      pattern
  End For
    Calculate the mean pattern overlap
    over all final states
Until the mean pattern overlap is less
than 95%
The Effective Capacity is P-1
```

3.2 Results

The experiments were implemented in a neural network with 500 and 1000 fully connected units (in previous experiments we found that the network size effects were insignificant providing the number of units was over 300). In previous studies it was found that a learning threshold of 10 gave a good performance of the network (Davey, et al, 2004). Thus for simplicity the learning threshold of the model is restricted to 10. This network was then trained with either bipolar or binary patterns, whose biases were varied from 0.1 to 0.9, and the EC values were measured. Each experiment was repeated 5 times and the average value together with the 95% confidence interval are reported.

Figures 2 and 3 give the main results of the experiments. In a previous study it was shown that the bipolar and binary networks perform the same when trained with unbiased patterns (Davey, et al, 2004). This result is confirmed here by the identical performance when the bias of the training set is 0.5. The performance of the bipolar and binary networks is significantly different when trained with biased patterns. With the bipolar representation, the performance is symmetrical about bias 0.5. That is,

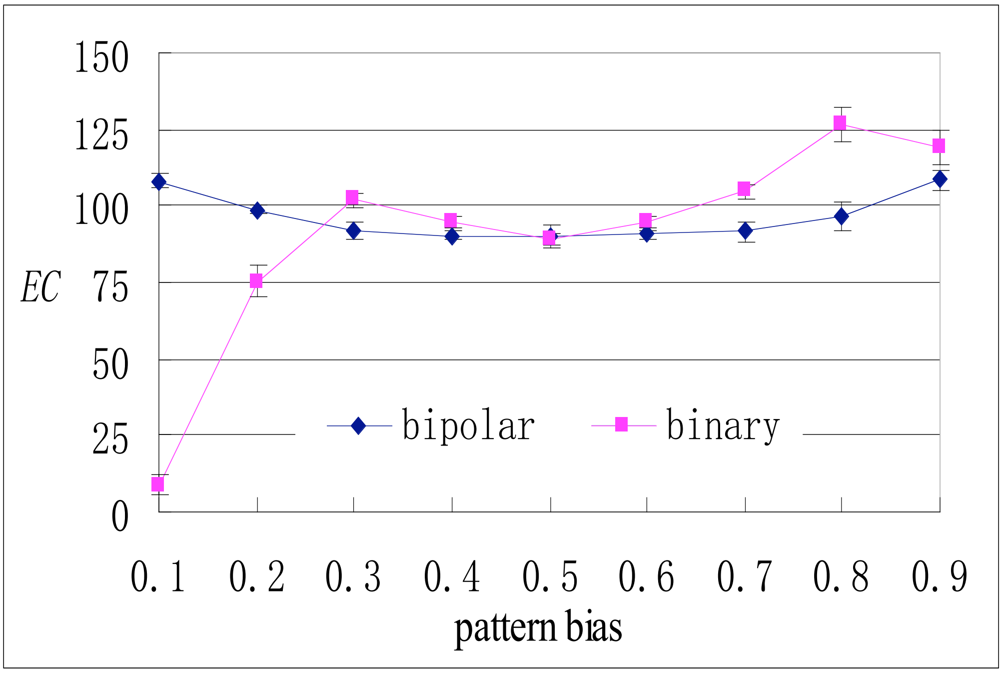

Figure 2. Effective Capacity results for a 500 unit, fully connected network with bipolar and binary representations. Biases of the patterns (as in the proportion of units which are on) are varied from 0.1 to 0.9. The results are averaged over 5 runs and intervals with 95% confidence are also given. The performance of the bipolar and binary network is identical when trained with unbiased patterns (ie bias = 0.5). With biased patterns, the binary representation performs better than the bipolar one, except for patterns of very low bias. The fall of performance of the binary network

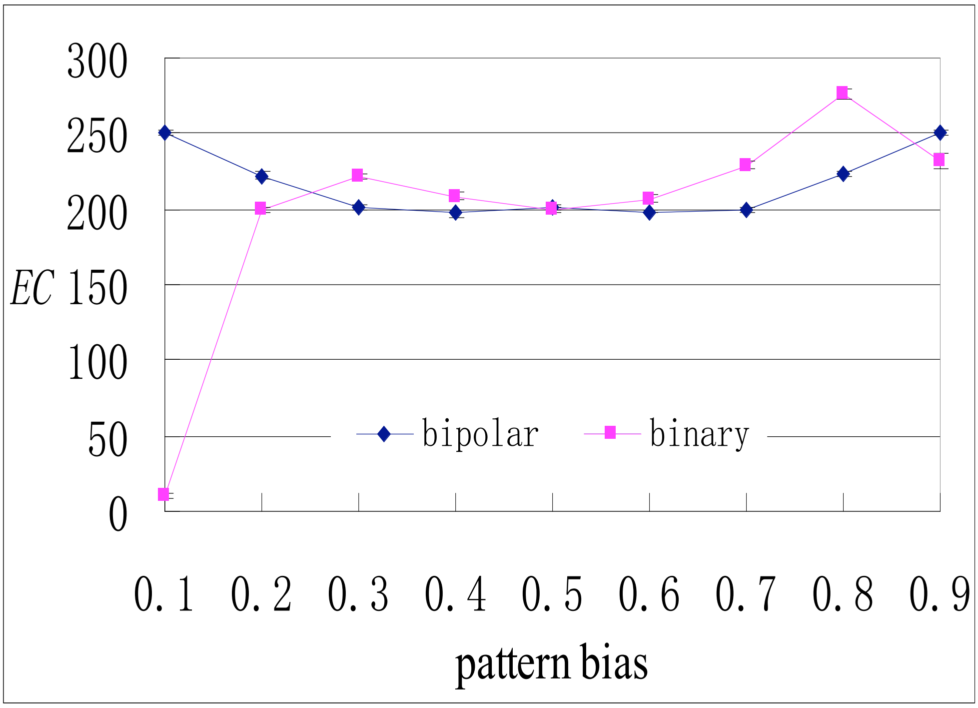

Figure 3. *Effective Capacity* results for a 1000 unit, fully connected network with bipolar and binary representations. Other settings are the same as Fig. 1. Results are similar to the 500 unit network.

for example, the EC at pattern bias 0.9 is identical to the one at pattern bias 0.1. This is of course a simple consequence of the symmetry of +1/-1. The result also indicates that the network performance is improved as the patterns become correlated. This is in line with Gardner's theoretical prediction (Gardner, 1988).

The results for the binary network are surprising. The first point to be made is that for most of the biases, the binary network performs better than or at least as well as the bipolar network. Only at the extreme of very low bias is the binary network significantly worse than the bipolar network. This is presumably due to the low proportion of units which are on. However, a detailed analysis of the binary network with training set bias of 0.1 finds that about 15% of the connections have no contribution to the network (the weights of these connections are zero), suggesting that the removal of these useless connections will improve the network's efficiency.

In the binary network, the performance falls when the bias is raised to 0.9. A detailed investigation indicates that it is caused by the significantly high attraction of the all 1 state, which is also found in the biased situation of a sign-constrained, bipolar network (Wong, 1992).

4 Conclusion

This paper extends Gardner's original model which used bipolar representation to a model with either bipolar or binary representation, and provides experimental results of their performances. The major finding of this paper is that although the performance of the binary representation is poor in the standard Hopfield network, it usually performs significantly better than the bipolar representation in a high capacity associative memory model trained with biased patterns. Only in the extreme situation where the bias of the training set is very low, does the binary representation performs worse that the bipolar one. These results are interesting since the binary and correlated patterns are more biological plausible than the bipolar, unbiased patterns which used in the traditional model.

Of course the real mammalian cortex is not a simple fully connected, binary network. In fact, researches on the connectivity of the mammalian cortex found that it is a so sparse network with special connecting strategies (Braitenberg and Schüz, 1998; sporns, et al, 2004). For example, the human's cerebral cortex has approximate 10^{11} neurons and 10^{14} connections, which means that each neuron is connected with only thousands of other neurons. So it is also interesting to investigate other aspects of the associative memory such as the connectivity effects (Davey, et al, 2006).

References

Abbott, L.F., Learning in neural network memories. Network: Computational Neural Systems, 1990. 1: p. 105-122.

Abeles, M., Studies of Brain Function. 1982, New York: Springer-Verlag.

Amit, D.J., Modeling Brain Function: The world of attractor neural networks. 1989, Cambridge: Cambridge University Press. 504.

Braitenberg, V. & Schüz, A. Cortex: Statistics and Geometry of Neuronal Connectivity, 1998, Berlin, Springer-Verlag.

Calcraft, L., Measuring the Performance of Associative Memories, in Computer Science Technical Report. 2005, University of Hertfordshire.

Calcraft, L., R. Adams, and N. Davey, Locally-Connected and Small-World Associative Memories in Large Networks. Neural Information Processing - Letters and Reviews, 2006. 10(2).

Davey, N., et al., High Capacity Associative Memory Models–Binary and Bipolar Representation. Proceedings of ASC 2004.

Davey, N., S.P. Hunt, and R.G. Adams, High capacity recurrent associative memories. Neurocomputing, 2004. 62: p. 459-491.

Davey, N., Calcraft, L, and Adams, R, High capacity, small world associative memory models. Connection Science, Volume 18, Issue 3 September 2006 , pages 247 – 264.

Gardner, E., The space of interactions in neural network models. Journal of Physics A, 1988. 21: p. 257-270.

Hopfield, J.J., Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences of the United States of America - Biological Sciences, 1982. 79: p. 2554-2558.

Wong, K.Y.M, and C. Campbell. Competitive attraction in neural networks with sign-constrained weights. Journal of Physics A: mathematical and General, 1992. 25, 2227.

Sporns, O. and J. D. Zwi, 2004. The small world of the cerebral cortex. Neuroinformatics 2(2): 145-62.