AN ASSOCIATIVE MEMORY MODEL WITH PROBABILISTIC SYNAPTIC TRANSMISSION

Bruce GRAHAM and David WILLSHAW
Centre for Cognitive Science, University of Edinburgh,
2 Buccleuch Place, Edinburgh, U.K., EH8 9LW

Abstract

The associative net model of heteroassociative memory with binary-valued synapses has been extended to include recent experimental data that indicates that in the hippocampus one form of synaptic modification is a change in the probability of synaptic transmission [2]. Pattern pairs are stored in the net by a version of the Hebbian learning rule that changes the probability of transmission at synapses where the presynaptic and postsynaptic units are simultaneously active from a low, *base* value to a high, *modified* value. Numerical calculations of the expected recall response have been used to assess the performance for different values of the *base* and *modified* probabilities. If there is a cost incurred with generating the difference between these probabilities, then the optimal difference is around 0.4. Performance can be greatly enhanced by using multiple cue presentations during recall.

1 Introduction

Information is stored in neural net models of associative memory by alteration of the strengths of the synaptic connections between units according to a local learning rule. These synaptic strengths are used to weight the input signals during pattern recall. Analysis of this type of model is relevant to neurobiological research, where long-term changes in synaptic strength have been suggested as the basis for learning and memory in the brain [1]. The exact nature of these changes is not completely understood and there may well be different synaptic learning rules in different parts of the brain. Recent data suggest that in the hippocampus one form of synaptic modification is a change in the probability of synaptic transmission [2]. We show here how the associative net can be used with this form of synaptic modification to give a stochastic net.

The associative net model of heteroassociative memory [3] consists of a layer of input units connected to a layer of output units by feedforward connections.

In the standard, deterministic net pairs of patterns with binary valued components are stored using a *clipped Hebbian* learning rule whereby the strength of the synaptic connection between an input unit and an output unit that are both active for the same pattern pair is modified from 0 to 1. During recall, a previously stored input pattern is placed on the input units as a cue. An output unit will receive one unit of activation from each active input unit to which it is connected by a modified synapse. A threshold-setting strategy (such as *winners-take-all*) is used to set the activity of the output units on the basis of their summed input activations.

The associative net is unique amongst associative neural network models in having binary-valued synapses. These synapses can be interpreted in terms of probabilities of synaptic transmission. In the standard model, a weight of 0 corresponds to 0% success in transmission, whereas a weight of 1 corresponds to 100% success. In the stochastic net the *clipped Hebbian* learning rule still operates but now synaptic modification involves changing the probability of transmission from a *base* value to a *modified* value. For example, a *base* synapse could have a probability (P_b) of 0.2 of transmitting a signal from an input to an output unit, while the corresponding probability for a *modified* synapse (P_m) could be 0.8. During recall, the summed inputs to an output unit will be affected by the probabilistic transmission through the synapses. Here we investigate the effects of different values of P_b and P_m on the recall properties of this stochastic associative net.

2 Methods

We have carried out numerical calculations of pattern recall from a net with $N_A = 8000$ input units and $N_B = 1024$ output units. Pattern pairs consisting of input patterns each containing $M_A = 240$ active input units and output patterns each containing $M_B = 30$ active output units were stored in the net. Recall of an output pattern was achieved by using it's associated input pattern as a cue and a *winners-take-all (WTA)* thresholding strategy. This involves each output unit measuring the weighted sum of its inputs *(dendritic sum)*. A unit's *dendritic sum* is different each time a cue is presented due to the probabilistic transmission. The same threshold is applied to all output units. It is lowered until M_B output units have a *dendritic sum* that is greater than or equal to the threshold. These units are made active and constitute the recalled pattern.

Recall performance is measured by pattern overlap and net capacity. Overlap is the correlation between the recalled pattern and the original stored pattern (1=perfect recall). Capacity is the maximum number of pattern pairs that can be stored so that all output patterns can be recalled reliably (at most one unit in error). With deterministic transmission ($P_b = 0$ and $P_m = 1$), this net has a capacity of 3576 pattern pairs.

The main point of interest is the effect of different values of P_b and P_m on recall performance. In particular, the same difference in probabilities can be achieved for different values of P_b and P_m. Three schemes for setting the

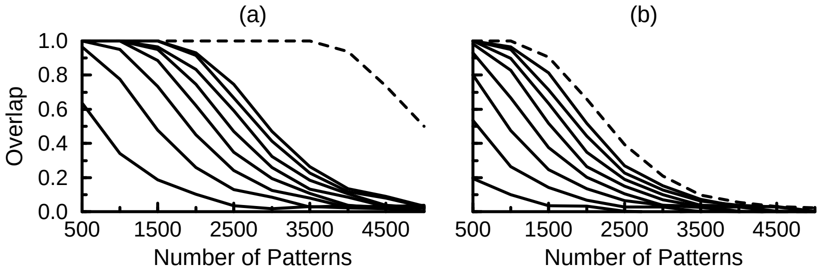

Figure 1: Recall overlap as a function of the number of stored pattern pairs for (a) a fully connected and (b) a partially connected net. The dashed line in each graph is the deterministic net. The solid lines have $P_m = 0.9$, with P_b from 0.1 to 0.8 in increments of 0.1, from right to left.

probability difference were investigated: *Low*: P_b varied, P_m fixed at 0.9; *High*: P_b fixed at 0.1, P_m varied; *Both*: both P_b and P_m varied, equidistant from 0.5.

3 Results

3.1 Recall overlap

When this stochastic associative net is fully connected, the only source of noise during recall, apart from the other stored patterns, is the probabilistic transmission at the synapses. Even for a large difference in probabilities (e.g. $P_b = 0.1$, $P_m = 0.9$), recall performance is considerably impaired compared to the deterministic net. Even so, many patterns can be stored and recalled without error. Examples of recall overlap when $P_m = 0.9$ and P_b is varied are shown in Figure 1.

Partial connectivity introduces another noise component during recall. The results shown here are when each output unit receives connections from a random selection of 60% of the input units. In this situation the performance of the deterministic net is degraded and is not significantly better than the stochastic net for high probability differences. This is shown by Figure 1(b).

3.2 Capacity

Figure 2 compares the capacity obtained at different probability differences, when the differences are set using the three schemes given above. For the fully connected net, consistently better performance is achieved when P_m is high (the *Low* scheme) (Fig. 2(a)). An optimal probability difference can be obtained by assuming there is a cost incurred that is proportional to the difference. The

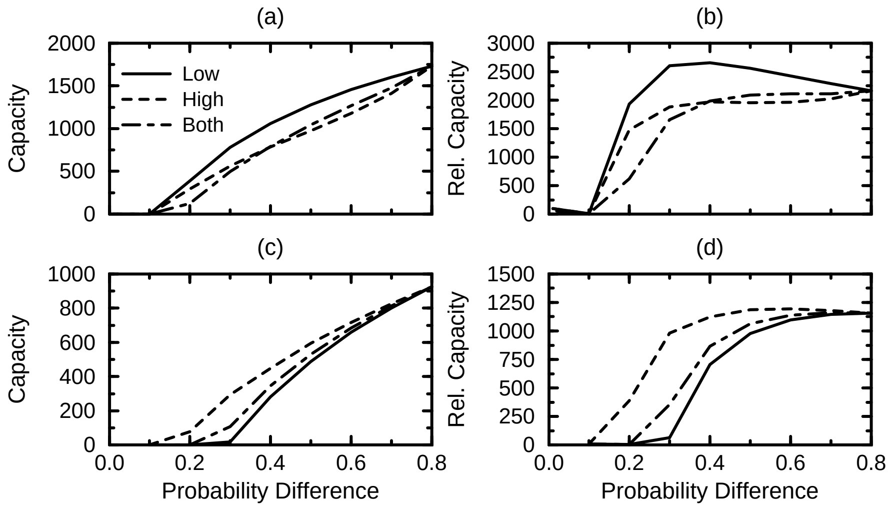

Figure 2: Capacity and relative capacity as functions of probability difference. (a),(b) Fully connected. (c),(d) Partially connected.

optimum can be found by considering the relative capacity, which is the net capacity divided by the probability difference. For the fully connected net using the *Low* scheme, the optimal difference is around 0.4 (Fig. 2(b)).

The capacity and relative capacity of the partially connected net are shown in Figure 2(c)&(d). Now the *High* scheme (low P_b) consistently provides a higher capacity than the other methods. The optimal probability difference is also now higher than for the fully connected net, being around 0.5 for the *High* scheme (Fig. 2(d)).

3.3 Multiple cueing

Performance can be improved by presenting the input cue many times so that each output unit can calculate an average *dendritic sum* before the WTA threshold is applied. Figure 3(a) shows the effect of multiple cueing on pattern overlap. Near optimum performance is obtained after 100 cue presentations.

The effect of multiple cueing on net capacity is shown in Figure 3(b). Capacity is highly dependent on the probability difference for only one cue presentation. The capacity for a probability difference of 0.8 (1729) is 4.5 times the capacity for a difference of 0.2 (387) after one cue presentation. However, after 1000 cue presentations the capacity at 0.8 (3204) is only 1.1 times the capacity at 0.2 (2964). Even a difference of 0.1 provides a large capacity if cues are presented 1000 times. If there is a cost associated with each cue presentation,

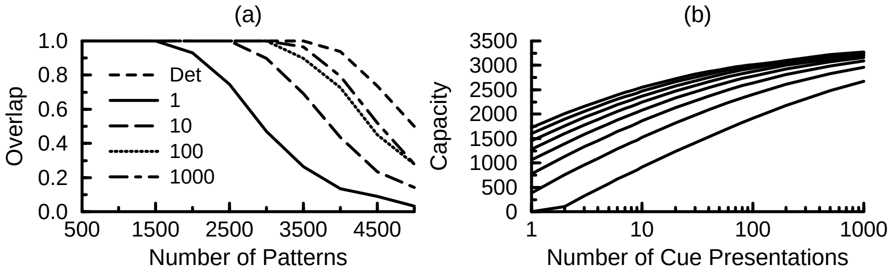

Figure 3: Multiple cueing with the fully connected net. (a) Recall overlap when the input cue is presented 1, 10, 100 or 1000 times (P_b = 0.1 and P_m = 0.9). (b) Capacity as a function of the number of cue presentations for the *Low* probability difference scheme (P_m = 0.9; P_b ranges from 0.1 to 0.8 going from top to bottom).

then five presentations is optimal for a difference of 0.1. A single presentation is optimal for greater differences.

3.4 Input Coding Rate

The effect of different input coding rates on net capacity is shown by Figure 4. Altering the input coding rate, M_A, alters the optimal probability difference. Higher values of M_A are less sensitive to the difference and so have a lower optimum. Lower values are more sensitive due to the smaller sample size these input patterns provide to output units during recall.

4 Summary

An associative net with probabilistic transmission functions well as a memory device. The optimum difference between the *base* and *modified* probabilities depends on net connectivity and pattern coding rate, but is generally small (≈ 0.4). Coincidentally, this corresponds to the magnitude of change seen experimentally [2]. Multiple cue presentation allows high performance with very small probability differences.

Acknowledgement: to the MRC for financial support under programme grant PG 9119632

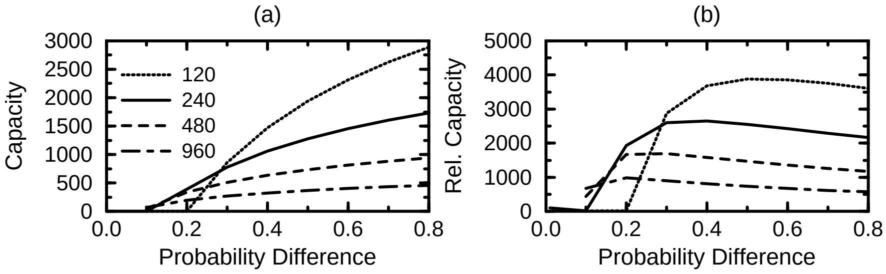

Figure 4: Capacity and relative capacity for one cue presentation to a fully connected net with input coding rates, M_A, of 120, 240, 480 and 960.

References

[1] Bliss, T.V.P. and Collingridge, G.L., *Nature*, 361:31-39, 1993.

[2] Stevens, C.F. and Wang, Y., *Nature*, 31:704-707, 1994.

[3] Willshaw, D.J., Buneman, O.P. and Longuet-Higgins, H.C., *Nature*, 222:960-962, 1969.