Brainwashing Random Asymmetric 'Neural' Networks

P. C. McGuire, G. C. Littlewort
J. Rafelski
Department of Physics, University of Arizona
Tucson, AZ 85721

31 May 1991

Abstract

An algorithm for synaptic modification (plasticity) is described by which a recurrently connected network of neuron-like units can organize itself to produce a sequence of activation states that does not repeat itself for a very long time. During the self-organization stage, the connections between the units undergo non-Hebbian modifications, which tend to decorrelate the activity of the units, thereby lengthening the period of the cyclic modes inherent in the network. It is shown that the periodicity of the activity rises exponentially with the amount of exposure to this plasticity algorithm. Threshold is also a critical parameter in determining cycle lengths, as is the rate of decay of the fields that accumulate at silent units.

Introduction

There are several potentially useful features in a randomly connected asymmetric network of quasi-neural elements [1] which acquires non-simple behavior of long cyclic modes. For example, the network explores many states and hence many associations without becoming trapped in a static state. A network without steady state attractors can invoke many different cyclic modes by applying different inputs or by global adjustments of neural firing thresholds, which allows access to a diversity of reverberations. Such a network is not dominated by a small number of final states with large basins of attraction, and so makes a suitable template for temporarily learning new modes.

It is not the aim of this paper to describe the applications of the system, but to develop a robust unsupervised learning procedure which produces networks with nearly chaotic behavior [2]. The network is chaotic, in the sense that when the network is in one very long cyclic mode, and a small amount of noise is introduced to the network for a very short time (causing a very few neurons to misfire), the network will go into a long transient, and sometimes enter into a vastly different long cyclic mode. The learning procedure is based on local information in space and time and does not depend critically on the choice of system parameters.

The Model

A network of binary threshold units is connected by a matrix of weights which undergo a period of "brainwashing" before the evolving activity of the units in the network is studied. The deterministic rules for updating the state of each unit, and for modifying the weights during self-organization, will be now described [3]:

- The update rule for the activity a_i (binary) of threshold unit 'i' at each time step depends on the accumulated local field c_i, the activation threshold V_i of the unit and the activation function, which is the step function $\Theta[x]$ for this model;

$$\begin{aligned} a_i(t+1) &= \Theta[c_i(t+1) - sV_i^0] && (1) \\ &= 0 \text{ or } 1 \\ c_i(t+1) &= \lambda c_i(t).(1 - a_i(t)) + \sum_j W_{ij}(t) a_j(t) && (2) \\ V_i^0 &= \frac{1}{2} \sum_k W_{ik} && (3) \end{aligned}$$

There is the usual network field arising from the activity of other units which connect into unit i via the weight matrix W_{ij}; and in addition the field is allowed to accumulate (decaying by factor $0 < \lambda < 1$ each time step) until the unit is activated, (firing, $a = 1$), at which stage the field is reset to zero. Normal thresholds, V_i^0 as defined in Eq. (3), are taken to be half the sum of the incoming weights to each unit, so that, on average, half the units with $a_i = 1$ (firing) will be just sufficient stimulus to reach threshold, provided there is no accumulated field. The actual threshold explored here is up to 20% different from this value, $0.8 \leq s \leq 1.2$.

- The weight matrix W is initially randomly chosen, such that there is a Gaussian distribution of sparse, asymmetric connection strengths (e.g. 60% of all connections are zero). The weight matrix is subsequently adjusted using the "brainwashing" algorithm for a short time, and is then constant during the process of searching for cycles in the network. The brainwashing procedure is of the form, for all i, j:

$$W_{ij}(t+1) = \eta W_{ij}(t)[1 - b a_i^{t+1} a_j^t] \qquad (4)$$

where b is the (small, positive) brainwashing rate parameter. The normalization constant η is introduced to ensure that the total 'strength' of the connection matrix remains unchanged:

$$\eta \sim 1/(1 - b\bar{a}^t \bar{a}^{t+1}) \qquad (5)$$

In other words, only weights connecting sequentially active units are modified, and the modification seeks to decorrelate the activity of the units it links, by making the relevant weights less effective. This rule is not simply an inverse of a Hebbian learning rule, because the direction of weight change is different for inhibitory connections, and changes are not additive but are proportional to the existing weight values, so weights do not change sign. The re-normalization indirectly increases the effectiveness of those weights not linking active units.

We denote the integrated brainwashing strength as B, i.e. the time T_{bw} the net is brainwashed for, multiplied by the increment b:

$$B = bT_{bw} \quad . \tag{6}$$

If the network activity is initialized many times with different $\vec{a}(t=0)$, subsequent to brainwashing, there should be several different cyclic modes (α) detected, with different cycle length (L_α), different sized basins of attraction (reflected in the frequency of occurrence of a particular mode ν_α) and different latency, or transient time before onset of periodic behavior.

Implementation

We have implemented and studied in detail the algorithm described above, for a network of 50 units. This is relatively small by network modeling standards, but because of the already extremely long cycles discovered by us in such small nets, the present work had to remain restricted. In the simulation results presented below, the random initial weights are chosen such that 30% of the weights are inhibitory and such that each unit has 20 incoming connections, giving a contact probability of 0.4.

The network then evolves according to the update rules Eqs.(1) to (3), with the decay parameter $\lambda = 0.5$ (i.e. if an element does not fire, half of the activity is retained until the next time step). We emphasize that $\lambda < 1$ makes network's non-firing elements at the time t remember the state of activity at time $t-1$. This so-called non-Markovian character of the network gives access to many more states than are available for a Markovian net ($\lambda = 1$), in which there is no memory retained about the synaptic condition of non-firing elements. The brainwashing parameter was taken b=0.02-0.05, when Eq.(4) is applied.

Another parameter of great interest is the magnitude of firing threshold. Normal thresholds, V_i^0 defined in Eq.(3) will cause elements to fire when the average activity is of the net is 0.5, provided there is no accumulated field. If fields do accumulate, then the average activity should be somewhat higher than 0.5, and this is indeed found (see below). If the thresholds are scaled globally by some factor s, such that $V_i^0 \to V_i = sV_i^0$ then lower thresholds lead to higher average activity and visa versa. By using different thresholds a whole new set of cyclic modes can be found.

Simulation Results

- **The effect of brainwashing:**
  The remarkable result presented in Figure 1 is that the average cycle length grows exponentially with the total brainwashing B. The usually small error bars imply

that cycle length can be rather reliably predicted, based on the system parameters and the integrated brainwashing B. We note that our simulations have yielded for the non-Markovian nets rather long cycles of rich structure.

- **The effect of firing threshold:**
  By using various different threshold levels s, a wide variety of cycles are found. Activity is found to decrease monotonically with threshold for both unbrainwashed (Figure 2) and brainwashed nets (Figure 3). For unbrainwashed nets, the average cycle length reaches a maximum at a threshold value which produces an average activity of 0.3 to 0.5 (Figure 4). It is thus possible, by adjusting the threshold level, to find long patterns of activity even in an unbrainwashed net. For slightly brainwashed nets, (Figure 5) the cycle lengths are much longer (note log scales), but the cycle length remains dependent on the threshold parameter, however, with a slightly broader peak. That means that in brainwashed nets the range of tolerable thresholds is slightly larger, and many more states become accessible through the device of shifting thresholds.

Latent Knowledge of the Net

We will next consider qualitatively if information has been stored 'accidentally' in the network. Clearly, this would be a naive explanation why many 'random' networks we and others have tried show normally rather uninteresting behavior. A good measure for this would be deviation from white noise of the correlation of activity measured at equal time. We thus compute the correlation matrix (changing notation from $a_i = 0, 1$ to $b_i = 2a_i - 1 = -1, 1$ for inactive and active units, respectively):

$$\begin{aligned} K^c_{i\neq j} &= \frac{1}{L_c}\sum_{t=1}^{L_c}(b^t_i - <b_i>_{L_c})(b^t_j - <b_j>_{L_c}) \\ &= <b_i b_j>_{L_c} - <b_i>_{L_c}<b_j>_{L_c} \end{aligned} \tag{7}$$

where L is the length of a typical cycle c used in the correlation calculation. In principle $|K_{ij}| < 2$, but if we drop the $i = j$ correlations, as done above, we expect to find a distribution strongly centered around zero. The result of our study are histograms of the elements K_{ij} shown in Figures 6-9 which were obtained for $[B, L] = [(0.0, 283), (0.1, 162), (0.2, 814), (0.3, 2825)]$. If the activity of each element were independent of others, such a histogram of 2,450 correlations would be close in form to a Gaussian, and naturally we should expect very long cycles in a non-Markovian system. The appearance of short recurrent cycles in unbrainwashed nets may be therefore associated with a significant 'shoulder' seen in Figure 6, which indicates that the elements are not fully independent in their activity - as is indicated by the large $\chi^2 = 5.5$ (per degree of freedom) of the Gaussian fit, (see the insert to Figures 6-9, for χ^2, normalization, location of the center of the Gaussian and its width parameter σ).

As we brainwash even a little, $B = 0.1$, we see in Figure 7 that this shoulder practically vanishes - instead the brainwashing algorithm generates a surplus near $K_{ij} = 0$, beyond and above what would be expected from a noisy network alone. In other words, we find that brainwashing is introducing a correlation which leads to the decorrelation of the individual units. The same result is seen for longer cycles and more brainwashing (Figures 8-9). Another interesting change is the reduction of χ^2 to a value O(2-3), which suggests that the Gaussian component in the correlation dominates the distribution.

Conclusions

The brainwashing algorithm was designed [1] to discourage steady states and short cycles, and to decorrelate the firing activity of the units in a recurrent network. We have demonstrated that it does this remarkably well, finding that the resultant networks have very long cycles which grow exponentially with the total amount of modification of the connections B. We also find that a rich variety of cyclic behavior is made accessible by small changes in the firing thresholds, and this variety grows with even a very small amount of brainwashing.

Seeking understanding why brainwashed nets show 'reliable' long cycles which we have first demonstrated in a non-Markovian system, we have considered the equal time activity correlation between units. Our expectation has been that the unsupervised learning algorithm somehow induces a strong pattern of correlations in a small portion of the network. We found, however, that while brainwashing removes accidental correlations inherent in the set up of the network, it generates strong anticorrelations between elements. We are not certain if this finding has the potential to explain how even well brainwashed nets are capable of finding cycling motion, despite their non-Markovian and nearly chaotic character. Presence of such regularity is suggestive of a significant order in the net, which we fail to detect in the above presented correlation analysis.

We have studied in this work in depth the brainwashing algorithm introduced earlier for the purpose of inducing volatility in 'dull' nets. We conclude that the brainwashing algorithm leads to extremely complex recurrent cycles, and imposes structures which are not totally random in the networks considered here.

Acknowledgements

We would like to thank David Harley and John W. Clark for help and interest in this work. P.C. McGuire would like to thank the Santa Fe Institute for hospitality and K.Atteson, S. Kauffman, G. Sonnenberg and CADMIUM for discussions and assistance.

References

1. Clark J.W., Rafelski J. and Winston J.V. (1985), Phys. Rep. 123, 215-273

2. Kürten K.E. (1988), Phys. Let. A129, 157-160

3. Shaw G.L., Silverman D.J. and Pearson J.C. (1985), Proc. Nat. Acad. Sci. 82,

2364-2368

FIGURE CAPTIONS

FIGURE 1:
Average cycle length L is shown to increase exponentially as a function of total brainwashing B. The curves for various values of rate b are shown.

FIGURE 2:
Average activity ($\alpha = \sum_{i,t} a_i^t$) is shown to decrease as a function of a threshold parameter s ($B = 0$).

FIGURE 3:
Same as Figure 2. but for a slightly brainwashed network ($B = 0.1$).

FIGURE 4:
Average cycle length $< L >$ is shown as a function of threshold scaling parameter s for an unbrainwashed network (B=0).

FIGURE 5:
Same as Figure 4. but for a slightly brainwashed network (B=0.1).

FIGURE 6:
Histogram distribution of *equal time* i, j correlation of firing activity K_{ij}, along with a Gaussian fit (B=0.0).

FIGURE 7:
Same as Figure 6. but for a slightly brainwashed network (B=0.1).

FIGURE 8:
Same as Figure 7. but for a slightly more brainwashed network (B=0.2).

FIGURE 9:
Same as Figure 8. but for a slightly more brainwashed network (B=0.3).