A Critical Analysis Of An Attractor Network Model Of Schizophrenia

Article · December 1996

DOI: 10.1007/978-1-4757-9800-5_74 · Source: CiteSeer

1 author:

Alastair G Reid
Oxford Health NHS Foundation Trust
21 PUBLICATIONS **152** CITATIONS

SEE PROFILE

A CRITICAL ANALYSIS OF AN ATTRACTOR NETWORK MODEL OF SCHIZOPHRENIA

Alastair Reid

al@cns.ed.ac.uk

Centre for Cognitive Science
University of Edinburgh, Scotland, EH8 9LW, UK

Abstract

Hoffman and Dobscha (1989) (H+D hereafter) used an attractor (Hopfield) neural network to model the effects of over-pruning in the human cerebral cortex in an attempt to demonstrate an underlying brain mechanism for schizophrenia. One of their main findings is the emergence of autonomous regions of activity in the network unrelated to input—they call these 'parasitic foci'. I have looked at the analytical work on attractor networks and show that parasitic foci are overlaps in spin glass states which automatically exist in the network, and that pruning is equivalent to raising the temperature in the stochastic state update equation. In addition I show that overloading rather than pruning the network will give the same qualitative results. This analysis is supported by various computer simulations. The conclusion is that while parasitic foci may form a very weak analogy with schizophrenic symptoms, their pathogenesis cannot be attributed to one particular process in the model. Even if the analogy holds the model has still not helped to elucidate the brain mechanism which underlies schizophrenia. I also discuss the inadequacy of the Hopfield network as a model of biological processes and schizophrenia.

Introduction

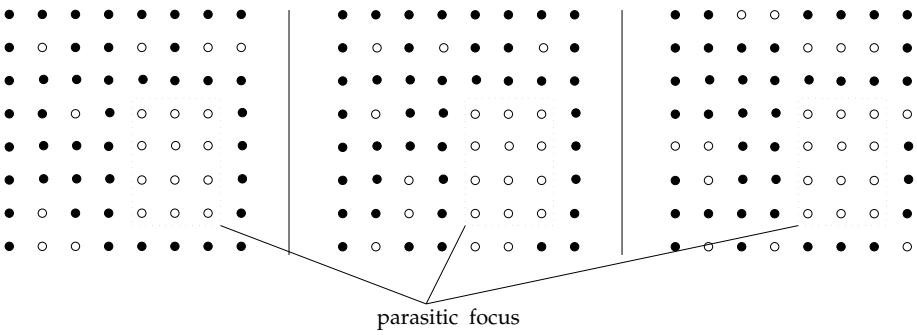
H+D hypothesised that pathological over-pruning in the frontal cortices gives rise to the positive symptoms of schizophrenia. To support this view they de-

vised a neural network model which when over-pruned gave various output patterns that were claimed to represent schizophrenic symptomatology. H+D used a 100-unit stochastic Hopfield-type network organised as a 10x10 grid and pruned according to the following rule:

$$\text{Prune connection from unit } i \text{ to unit } j \text{ if}$$
$$w_{ij} < (p \times distance\, between\, unit\, i\, and\, unit\, j)$$

p is the pruning coefficient and values ranged from 0.6 to 1.0. The network stored 9 memories. Inputs to the networks consisted of the stored memory patterns with either every fifth or every third bit flipped. This gave two sets of input patterns with respectively Hamming distances of 20 and 33 from the stored memories. These inputs were intended to represent two different levels of ambiguity for the memory model to cope with. Over-pruned networks showed the following three 'pathological' output states:

- **generalisations:** An amalgamation of memory fragments
- **loose associations:** Output with more than 10 bits different from input and not a generalisation
- **parasitic foci:** Certain populations of neurons converging on the same non-memory output regardless of input. These are determined by comparing all the loose association outputs for one run of a network. An area of overlap which is at least a block of 12 units in size (either 3x4 or 2x6 units) occurring between all the output states examined is said to be a parasitic focus. The figure below represents three loose association end-states of a network for three different inputs. The white dots represent units with the same activation across all three output patterns, black dots are units with different activations.



It was claimed that these network activities correspond to psychotic phenomena such as thought disorder and auditory hallucination.

I have replicated H+D's original pruning experiments. I extended my simulation to include:

1. unpruned networks
2. pruned networks without self-weighted terms (i.e. $w_{ii} = 0$)
3. randomly pruned networks
4. unpruned overloaded networks
5. pruned and unpruned networks storing 5 patterns

Simulation results are the average of 10 different networks for each degree of pruning or overloading. Pruning coefficients ranged from 0.6 to 1.05; the number of patterns stored in overloaded networks ranged from 10 to 19.

Results

Type of Network	No of parasitic foci
9-patterns pruned	24
9-patterns unpruned	31
Overloaded	13
5-patterns pruned	33
5-patterns unpruned	1

1. unpruned 9-pattern networks produced parasitic foci to the same extent as pruned networks.

2. the removal of self-weighted terms produced fewer loose associations but did not affect the number of parasitic foci.

3. randomly pruned networks functioned poorly as memories when the number of connections removed was the same as the number lost in the rule-based case (60-85% loss). They did not retrieve successfully any memories and produced no parasitic foci. When the number of connections lost randomly was lower (20-55% loss) the results obtained were similar to the rule-based case.

4. overloaded networks showed very similar behaviour to pruned networks but produced fewer parasitic foci.

Analysis

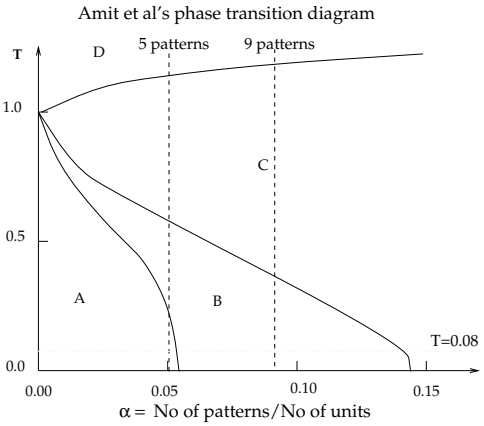
We can consider H+D's model in terms of the phase transition diagram for the stable states of a Hopfield network produced by Amit et al. (in Hertz, Krogh and Palmer (1991))This diagram shows the stability in the $T - \alpha$ plane of the states of a stochastic attractor network. In region A the desired memories are the most stable states, although mixture states are also stable. In region B the desired memories are also stable but spin glass states are more stable. In region C only the spin glass states are stable and in region D no states are stable.
In analytic terms H+D's *generalisations* are mixture states; *loose associations* are spin glass states; and *parasitic foci* are parts of spin glass state outputs from a

particular network which overlap with each other to a significant degree.

In a stochastic attractor network S_i, the state of unit i, is set by:

$$Prob(S_i = 1) = \frac{1}{1 + e^{(-\frac{2}{T}h_i)}} \tag{1}$$

Where T is the 'pseudo-temperature' and h_i the net input to unit i. The diagram represents phase transitions in terms of the *normalised* value of T and α. Normalised in the sense that it is usual to include a $\frac{1}{N}$ term in the equation for h_i, and the stochastic update equation used by H+D used $\frac{1}{T}$ rather than the usual $\frac{2}{T}$ in the exponential term. If T_{norm} is the usual (normalised) form of T and T_{hoff} the form of T used in their simulations then the following equivalence holds:



$$T_{norm} = \frac{2}{N} T_{hoff} \tag{2}$$

Thus the value $T = 4$ in H+D's model is equivalent to a normalised value of $T = 0.08$. **Pruning the network is equivalent to weak dilution.** Under this situation the relative concentration of connections c is given by:

$$c = 1 - (\frac{Number\ of\ connections\ removed}{Total\ number\ of\ connections}) \tag{3}$$

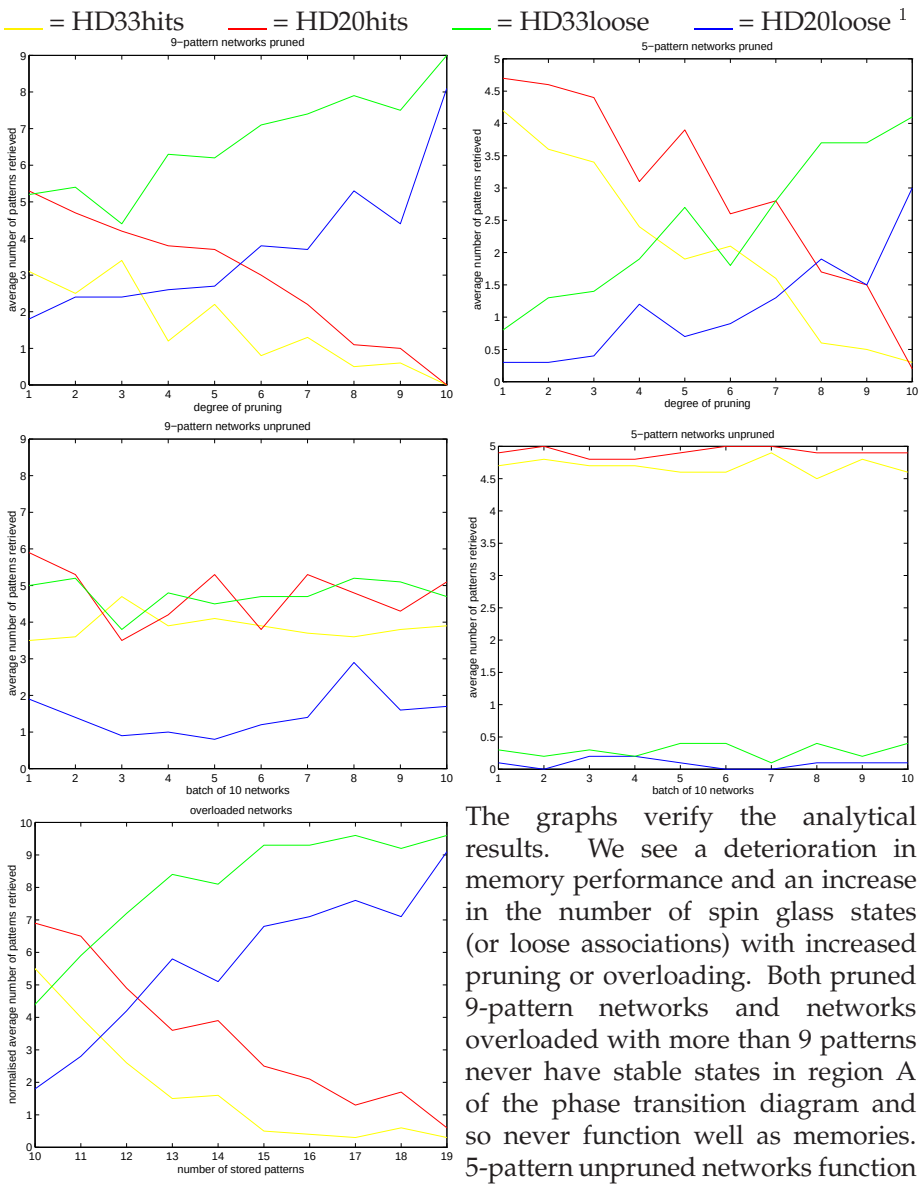
the input h_i in the pruned case then becomes:

$$h_i = c\Sigma w_{ij}^{unpruned} S_j \tag{4}$$

$$h_i = c h_i^{unpruned} \tag{5}$$

This is equivalent to increasing T in the state update equation by $\frac{1}{c}$. The effect of increasing T, in terms of phase transitions, is to move the stable states of the system away from region A and into regions B and C where spin glass states predominate. This accounts for the gradual deterioration of a network's performance with pruning. Similarly, however, if a network is overloaded then clearly the size of $\alpha$is increased and the effect is to move along the x-axis of the phase transition diagram. Again the stable states of the system move from region A to regions B and C

The performance of pruned and overloaded networks is shown in the graphs:



The graphs verify the analytical results. We see a deterioration in memory performance and an increase in the number of spin glass states (or loose associations) with increased pruning or overloading. Both pruned 9-pattern networks and networks overloaded with more than 9 patterns never have stable states in region A of the phase transition diagram and so never function well as memories. 5-pattern unpruned networks function as near perfect memories and 9-pattern unpruned networks function as moderate to poor memories, again as predicted by the phase transition diagram.

[1] HD33hits represents correct retrieval of memories or generalisations with inputs of hamming distance 33 away from the stored memories. HD33loose represents retrieval of loose associations. Likewise for HD20hits and HD20loose

Summary and Conclusions

There are problems with H+D's methodology and with their model.
• Specifically it is unusual to include self-weighted terms in a Hopfield net as these increase the number of spurious patterns.
• H+D have used 9-pattern networks which have parasitic foci even when unpruned. While 5-pattern networks show genuine transitions to pathological states when pruned, these states can also be obtained by overloading the network rather than pruning it. Thus *pruning is not the only cause of parasitic foci*. This seriously undermines the use of this model to show a potential brain mechanism for schizophrenia.
• The motivation for the pruning rule is unclear. Pruning and synapse formation are both *developmental processes* but synapse formation is not included in the model. The pruning rule uses fixed weights which implies that pruning takes place at an instant in time, and after the cognitive processes which the model represents have been formed i.e. in the mature brain. Pruning is neither instantaneous nor a mature process in real brains.
• Allowing connection strengths to be symmetric (i.e. $w_{ij} = w_{ji}$) is also a problem in terms of the biological realism of the model.
• Another problem is that such a net is essentially a model of auto-associative long-term memory (LTM). If schizophrenic symptomatology arises through the breakdown of LTM function then we should see deficits in LTM occurring with these symptoms. This is certainly not the case.
• The main fault is that the model *lacks biological realism* and this arises entirely from using a Hopfield net to model biological processes.

While parasitic foci are real phenomena in a Hopfield net it seems unlikely that they actually occur in this way in the brain. However, the idea that alterations in the nature of the attractors that exist in real brain dynamics could describe the pathogenesis of schizophrenic symptoms remains very exciting. With a greater attention to anatomic and physiologic detail a model could be constructed to investigate this in more detail and give us further insight into the mechanisms of this illness.

References

Hertz, J; Krogh, A; Palmer, RG (1991): *Introduction to the theory of neural computation*. Pubs Addison Wesley

Hoffman, RE; Dobscha, SK (1989): Cortical Pruning and the Development of Schizophrenia: A Computer Model. *Schizophrenia Bulletin* Vol 15 No 3 pp477-489