Neural networks with transparent memory

1988 J. Phys. A: Math. Gen. 21 L259

(http://iopscience.iop.org/0305-4470/21/4/013)

J. Phys. A: Math. Gen. 21 (1988) L259-L262. Printed in the UK

LETTER TO THE EDITOR

Neural networks with transparent memory

J F Fontanari and R Köberle

Departmento de Física e Ciência dos Materiais, Instituto de Física e Química de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560 São Carlos SP, Brazil

Received 3 December 1987

Abstract. Neural networks are only useful as associative memories if an outside observer is able to ascertain whether a particular input corresponds to one of the memory states or not, i.e. the memory must be transparent. We design a model system which enters a 2-cycle whenever an input is not within the basin of attraction of some memory state.

Neural networks have recently attracted a lot of attention since they are the best way known to implement non-programmable, massively parallel and robust computing systems with associative memories.

A very simplified model would consist of N neurons described by variables S_i, $i = 1, 2, \ldots, N$, which may assume the values $+1$ for active and -1 for passive neurons and synaptic connections J_{ij} encoding the strength with which neuron j acts on neuron i.

In Hopfield's model (Hopfield 1982) J_{ij} are symmetric and determined by the patterns of activity to be stored as memories. If we call these patterns prototypes $\xi_i = \pm 1$ then the storage of p of them is achieved by implementing Hebb's rule as

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu}. \tag{1}$$

In this model, update is asynchronous and can be deterministic or random with the presence of some noise. These two cases are referred to as the zero-temperature limit and finite temperature. We will restrict ourselves to the former in the following for simplicity.

If N is large and the prototypes are orthogonal, the number of prototypes which can be stored is proportional to N:

$$p = \alpha N. \tag{2}$$

Numerical simulations (Amit *et al* 1987) show that for $0 \leqslant \alpha \leqslant \alpha_c \approx 0.14$ an input pattern sufficiently near to one of the prototypes will lead the system to a fixed point which is very close to the prototype. This proximity is measured by the retrieval overlap

$$m^{\mu}(t) = \frac{1}{N} \sum_i \xi_i^{\mu} S_i(t) \qquad \mu = 1, \ldots, P \tag{3}$$

and for $\alpha \leqslant \alpha_c$ it yields $1 \geqslant m^{\mu}\ (t = \infty) \gtrsim 0.97$. The system is in the retrieval phase. If $\alpha_c < \alpha < \infty$ the system will still reach a fixed point but with $0.4 \gtrsim m^{\mu} \gtrsim 0.18$. The system is then said to be confused.

As has been pointed out recently by Parisi (1986) such a system has an important shortcoming which renders it useless as a realistic, usable associative memory. The problem is the following. When some input pattern is presented and we want, for example, to ask the simple question whether this pattern is one (or near one) of the prototypes, it is impossible to get an answer, since without previous knowledge of the memory contents we do not know whether the stationary state attained corresponds to a prototype or whether the system is confused. An outside observer would simply be unable to discriminate between the two cases.

In a workable system it would be extremely useful if these two situations could be distinguished. Parisi has suggested that such a discrimination should be possible for sufficiently asymmetric synaptic connections, but this has not been verified.

The above mentioned problem can be explicitly solved if we resort to synchronous update. Models with synchronous (parallel) dynamics have recently been investigated in some detail and found to exhibit a very rich phase diagram (Fontanari and Köberle 1987, 1988). The new feature distinguishing these systems is the presence of cycles, at most of length two, which can never occur for asynchronous models. Neurons participating in a cycle flip from $S_i(t)$ to $S_i(t+1)=-S_i(t)$. Therefore if we want cycles to occur in a controllable fashion, our model should include the possibility of this switching behaviour being a self-regulating process. In our previous studies we found that the intensity of the diagonal coupling J_{ii} is a parameter whose tuning allows the occurrence of cycles to be controlled. Now we want the system itself to do the tuning.

The idea then would be to design a system which goes to a fixed point if the retrieval overlap, equation (3), is greater than a given parameter $\bar{m}$. Accordingly we will include in the model a diagonal coupling which is small for $m^\mu \geqslant \bar{m}$ and large and negative (it is negative values of J_{ii} which favour cycles) for $m^\mu < \bar{m}$.

With this in mind we propose the following synchronous evolution:

$$S_i(t+1)=\operatorname{sgn}[h_i(t)] \tag{4a}$$

where

$$h_i(t)=\sum_j J_{ij}(t)S_j(t)$$

and $J_{ij}(t)$ is determined as

$$J_{ij}(t)=\begin{cases}\dfrac{1}{N}\displaystyle\sum_{\mu}^{p}\xi_i^\mu\xi_j^\mu & i\neq j\\[2ex] -p\displaystyle\prod_\mu H\left(\bar{m}-\frac{1}{N}\sum_l \xi_l^\mu S_l(t)\right) & i=j\end{cases} \tag{4b}$$

$H(x)$ being the step function

$$H(x)=\begin{cases}1 & x>0\\ 0 & x\leqslant 0.\end{cases}$$

Thus, given the initial configuration $\{S_i(t_0)\}$ we use (4b) to compute $J_{ij}(t_0)$, which in turn inserted in (4a) yields $\{S_i(t_0+1)\}$. Notice that the off-diagonal part of J_{ij} is time independent.

If $m^\nu(t)$ is greater than $\bar{m}$ the term $\mu=\nu$ in the product for $J_{ii}(t)$, equation (4b), will vanish and only the usual off-diagonal part survives, leading the system to a fixed point. If $m^\nu(t)$ is smaller than $\bar{m}$ for all ν then $J_{ii}=-p$ will overwhelm the off-diagonal parts. Thus if the system enters a limit cycle, the input pattern is not near a prototype.

The value of $\bar{m}$ determines the size of the basins of attraction of the retrieval states and any value greater than $\langle m^{\mu}(\alpha=0.15)\rangle_{\xi}$, where $\langle\ldots\rangle_{\xi}$ means an average over prototypes, will not modify the system's performance described above. Although it is known that $\langle m^{\mu}(\alpha=0.15)\rangle_{\xi}\approx 0.4$ for Hopfield's model (Amit *et al* 1987), there is no similar result for the synchronous model and it is beyond our computational resources to calculate it. However, we expect $\langle m^{\mu}(\alpha=0.15)\rangle_{\xi}\rangle$ 0.4 for the synchronous model (Gardner *et al* 1987).

In order to test if our model performs the way described we measure the average 'magnetisation'

$$m=\langle(m^{\mu}(t)+m^{\mu}(t+1))/2\rangle_{\xi} \tag{5}$$

and

$$q=\left\langle\frac{1}{N}\sum_{i} S_i(t)S_i(t+1)\right\rangle_{\xi} \tag{6}$$

where $(1-q)/2$ yields the probability of finding complete cycles (all spins flipping) and $(1+q)/2$ the one of finding fixed points.

We ran simulations for $\bar{m}=0.97$ and $N=50$, 100 and 200, averaging over 3000 prototypes. The input patterns were chosen to be prototypes. As can be seen from figure 1(a), the magnetisation drops from its maximum value at $\alpha=0$ rather sharply to zero when α varies through $\alpha_c\approx 0.14$. We obtained $m=0$ for $\alpha>\alpha_c$ because $m_{\mu}(t)=-m^{\mu}(t+1)\neq 0$. For small values of m the system enters a cycle. This is shown in figure 1(b), where we show q as a function of α. The region around α_c where fixed points ($q=+1$) coexists with complete cycles ($q=-1$) is expected to disappear in the large-N limit. As mentioned above, these results are in fact independent of $\bar{m}>\langle m(\alpha=0.15)\rangle_{\xi}$. Notice that there will be no fixed points near $\{-\xi_i^{\mu}\}$ since $J_{ij}(t)$ is not invariant under the replacement of $\{\xi_i^{\mu}\}$ by $\{-\xi_i^{\mu}\}$.

Summarising, this letter presents a model whose memory contents can be known by the external world, i.e. they are transparent, since the retrieval and confused phases have different asymptotic dynamical behaviours.

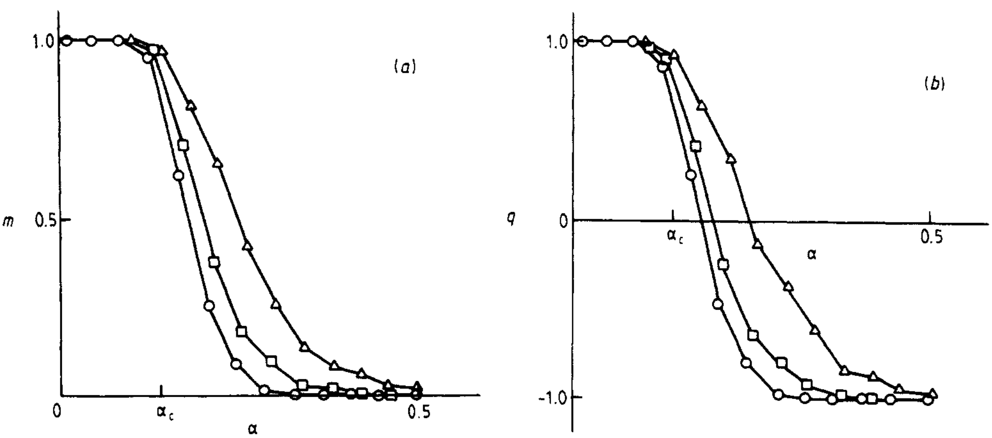

Figure 1. (a) Average retrieval overlap for $\bar{m}=0.97$ and $N=50$ (△), 100 (□), 200 (○) as a function of α. We average over 3000 prototypes which are chosen to be the input patterns. (b) Average overlap between equilibrium states for consecutive times. The parameters are the same as those of (a).

The research of RK is partially supported by CNPq and JFF holds a FAPESP fellowship. Simulations were carried out on a VAX 780/11 computer.

Note added in proof. The authors thank the referee for pointing out the work of Shinomoto (1987) and Skarda and Freeman (1986) where ideas related to our work have been formulated.

References

Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys., NY* **173** 30
Fontanari J F and Köberle R 1987 *Phys. Rev.* A **36** 2475
—— 1988 *J. Physique* in press
Gardner E, Derrida B and Mottishaw 1987 *J. Physique* **48** 741
Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
Parisi G 1986 *J. Phys. A: Math. Gen.* **19** L675
Shinomoto S 1987 *Biol. Cybern.* **57** 197
Skarda C A and Freeman W J 1986 *Behavioral and Brain Sciences* (Cambridge: Cambridge University Press)