VALIDATION IN DISTRIBUTED REPRESENTATIONS

S H Srinivasan   M Narasimha Murty
Department of Computer Science & Automation
Indian Institute of Science
Bangalore - 560 012, **INDIA**

Abstract

In the Hopfield model of content addressable memory, the number of spurious attractors is exponential in the dimensionality of the memory. Hence it is highly 'likely that the system converges to a spurious memory on an arbitrary input. It is desirable that the system has a way of checking whether its state corresponds to any one of the stored patterns. In this paper we show that it is possible to validate the patterns in a distributed fashion. The model uses complex activations and synapses for this purpose. The genuine memories **are** defined to be real attractors. Under some assumptions we show that the expected number **of** spurious attractors is negligibly small. **We** also calculate the capacity of the model.

One of the characteristic features of human memory is content addressability. Based on partial input cues, the system retrieves one of the stored patterns. Implicit in this statement is the fact that the memory "knows" that the retrieved item is something it has encountered before. People fail to remember but they seldom confuse unknown things with known things (unless they are hallucinating). Thus we need models of memory which, in addition to having content, addressability, can validate their input in light of their experience. The problem of validating the internal states of the system is called the *validity* problem.

To make matters more concrete, consider the Hopfield model of content addressable memory[1]. Two important considerations of any neural model are: learning and computation. In this case these two correspond to storage and retrieval. In the Hopfield network, Hebb rule (also called the sum of outer products, in this case) is used for learning; relaxation is used **as** the mode of computation. During recall, the network iterates in its state space, $\{\pm 1\}^N$, until convergence is achieved. The final state is taken **as** the recalled memory item. It is possible to show that the relaxation converges in all cases if the units update themselves asynchronously. This can be proved by showing the existence of a function whose value decreases after each asynchronous update. Hence we can consider the stable states of the model as the local minima of this function. For this reason, the memories are also called the "attractors".

In the context of dynamical systems, it is desirable to distinguish between *equilibrium points* and *stable equilibrium points* or *attractors*. An upper bound for the number of equilibrium points of the Hopfield network is N [2],[3]. What is desirable for a CAM is that the memories should also be stable equilibrium points. The maximum number of memories that can be stored **as** stable attractors is called the *capacity* of the model. If Hebb rule is used for storage then the capacity of the Hopfield model is asymptotically $\frac{N}{2\log N}$ [4]. Another important performance measure of a CAM is the number of spurious memories (equilibrium points or attractors). The number of equilibrium points of the Hopfield network is exponential in the dimensionality [4][5]. Since at most N arbitrary patterns can be made equilibrium points of the model, the number of spurious equilibria is exponential in N. Hence there is a great likelihood that the network converges to a spurious attractor on an

arbitrary input. The system **has** no way of checking whether the attractor corresponds to one of the stored memories. We need a mechanism to validate the state of the network in light of experience.

In this paper we propose a mechanism to solve this problem. The main idea is to make the network **use** complex numbers for its operation. In the complex domain, we define the genuine memories as *real* stable equilibrium points. We can then expect that the number of spurious memories reduces greatly. In the rest of the paper we show that the above ideas indeed work.

1   The model

As already mentioned we want to introduce complex numbers to the basic model. One way to do this is to introduce an output function and make this function complex. In the usual Hopfield model the output is the identity function; in other words, the output of a unit is the same **as** its activation. In our model, the output of a unit is given by $f(S)$, where S is the activation of the unit. Introducing an output function raises the following issue: since the activation vector and the output vector will in general be different, which should be taken as the state of the network? Put differently, after the dynamics of the network converges, does the activation vector or the output vector stand for the recalled item? In our model the state of the network is the activation vector.

The learning rule is a modification of the sum of outer product rule. If $S_1, S_2, \cdots, S_M$ are the patterns to be stored, where $S_\mu \in$ then the strength of the connection from unit k to unit i is given by

$$
\begin{aligned}
w_{ik} &= u_{ik} + \hat{j} v_{ik} \\
&= \sum_{\mu=1}^{M} S_i^\mu \overline{f(S_k^\mu)}
\end{aligned} \tag{1}
$$

where $\overline{z}$ denotes the complex conjugate of z and $\hat{j} = \sqrt{-1}$. We set $w_{ii} = 0$. The connection strengths are asymmetric, in general.

Let the current state of the network be $\mathbf{S} = (S_1\ S_2 \cdots\ S_N)^T$. The operation of a unit can be summarized as

1. calculate the net input, F_i, as

$$
\begin{aligned}
F_i &= G_i + \hat{j} H_i \\
&= \sum_{k=1}^{N} w_{ik} f(S_k)
\end{aligned} \tag{2}
$$

2. update the activation value to S_i' as $S_i' = \Theta(F_i)$. where $\Theta(\cdot)$ is a complex hard-limiting function.

3. output $f(S_i')$.

The updates can be synchronous or asynchronous. The following subsections discuss the choice of functions $\Theta(\cdot)$ and $f(\cdot)$.
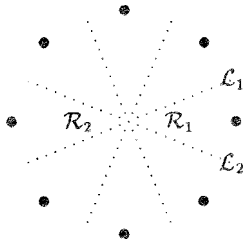
Figure 1: Complex hard-limiting function. The figure shows the complex plane. The receptive field for each activation is the region surrounding it. For example, if the net input in region $\mathcal{R}_1$ then the activation value of she unit is updated *to* +1. Similar comments hold for region $\mathcal{R}_2$ and activation value --I. The equation of line $\mathcal{L}_1$ is $y = m x$ where $m = \tan(\pi/8)$; that of $\mathcal{L}_2$ is $y = -mx$.

1.1 Complex hard-limiting function

This function relates the net input and the activation of a unit. We use a finite number of activation d u e s. We also make the convention that all the activation values have unit, magnitude, i.e., they lie on the unit circle in the complex plane. The most natural way to define the activation function is to divide the complex piane in to disjoint regions and assign one activation value for each region. The updated activation value of a unit is the activation value corresponding to the region in which the net input falls. These regions are also called "receptive fields" of the activation values. For the analysis in this paper it is sufficient to define the receptive fields of +1 and −1. But as an example, we define $\Theta(\cdot)$ as below: *(see figure 1)*

$$\Theta(\rho e^{j\theta}) = \begin{cases} 1, & \frac{-\pi}{8} \leq \theta < \frac{\pi}{8} \\ e^{j\frac{\pi}{4}}, & \frac{\pi}{8} \leq \theta < \frac{3\pi}{8} \\ j, & \frac{3\pi}{8} \leq \theta < \frac{5\pi}{8} \\ e^{j3\frac{\pi}{4}}, & \frac{5\pi}{8} \leq \theta < \frac{7\pi}{8} \\ -1, & \frac{7\pi}{8} \leq \theta < \frac{9\pi}{8} \\ e^{j\frac{5\pi}{4}}, & \frac{9\pi}{8} \leq \theta < \frac{11\pi}{8} \\ -j, & \frac{11\pi}{8} \leq \theta < \frac{13\pi}{8} \\ e^{-j\frac{\pi}{4}}, & \frac{-3\pi}{8} \leq \theta < \frac{-\pi}{8} \end{cases}$$

where $\rho e^{j\theta}$ is a complex number with magnitude ρ and argument θ. This function uses 8 activation values spaced uniformly on the unit circle in $\mathbb{C}$.

1.2 Output function

This function relates the activation of a unit to its output. If S is the state of the network then the net input to a unit is given by

$$F_i = \sum_\mu S_i^\mu \sum_k \overline{f(S_k^\mu)} f(S_k) \qquad (3)$$

We get this by substituting equation 1 in 2. We will be interested in this sum for real S. If we assume that $f(\cdot)$ also takes values on the unit circle in $\mathbb{C}$, then it is easy to see that F_i is the same for all the output functions for which the angle, ϕ, between $f(+1)$ and $f(-1)$ is the same. Hence the results in the subsequent sections are given in terms of ϕ. The output function can be taken as $f(1) = 1$ and $f(-1) = e^{j\phi}$.

2 Number of spurious memories

Given the above details we proceed to calculate the number of spurious memories. If we assume that the vectors to be stored are independent random variables with equal probability of ±1, then the connection strengths in equation 1 **are sums** of independent random variables, and hence are normal. Consider a random binary input vector which may or may not be one of the stored patterns. Without loss of generality we can consider this *to* be composed of
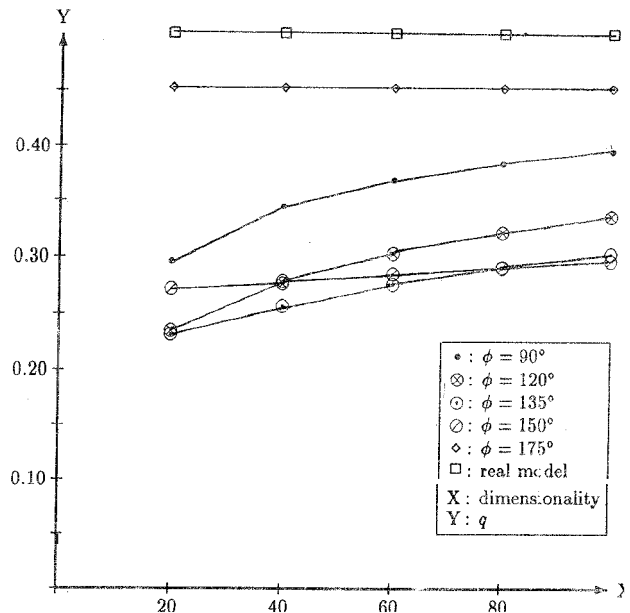
Figure 2: variation of q with angle and dimensionality

+1s. The probability, p, that this is an equilibrium point of the network is given by

$$p = \text{Prob}\{F_1 \in \mathcal{R}_1, F_2 \in \mathcal{R}_1, \cdots, F_N \in \mathcal{R}_1\}$$

and the expected number of real equilibrium points of the network by $p2^N$. This section concentrates on the calculation of p.

2.1 A quick calculation

In this we assume that the weights u_{ik} and v_{jl} are independent $\mathcal{N}(0,1)$ random variables. This is equivalent to the assumption under which the expected number of spurious equilibrium points of the real Hopfield model is calculated to be $(1.0505)2^{0.2874N}$ [4][5]. Since the network is asymmetric, the net inputs F_i, at different. units are independent. Hence

$$p = q^N, \quad \text{where} \quad q = \text{Prob}\{F_i \in \mathcal{R}_1\}$$

Letting $F_i = G_i + jH_i$, we can calculate that both G_i and H_i are $\mathcal{N}(0, (N-1))$ and are uncorrelated (and hence independent). Then

$$q = \int\int_{\mathcal{R}_1} n(x, 0, (N-1); y, 0, (N-1); 0) \, dx \, dy$$

where $n(x, \mu_x, \sigma_x^2; y, \mu_y, \sigma_y^2; p)$ is the bivariate Gaussian density function with means μ_x and μ_y, and variances σ_x^2 and σ_y^2 with correlation coefficient p. Because of symmetry of the integrand, the value of the integral is the fraction of the area of the complex plane enclosed by $\mathcal{R}_1$, which is $\frac{1}{8}$ (see figure 1). So $p = \frac{1}{8^N}$ and the expected number of spurious equilibria is $2^N p = \frac{1}{4^N}$.

The above calculation is confirmed by simulations. A network with $N = 16$ and $\phi = 150°$ for the output function was simulated. Both the real and imaginary parts of the weights were $\mathcal{N}(0,1)$. Each of the 2^{16} binary vectors was checked for equilibrium condition None of the vectors were equilibrium points of the network.

2.2 Realistic calculation

The above calculation assumes random asymmetry But weights learned using equation 1 have functional asymmetry

163

asymptotic calculation of the expected number of spurious equilibria is beset with technical difficulties mentioned in 4] i.e., we need to consider an increasing number of weights. Since exact calculation seems impossible, we indicate below why the expected number of spurious equilibria should be small for some values of ϕ.

If we use equation 1 to calculate the weights, the weights are related in the following way:

1. u_{ik} and v_{ik} are $\mathcal{N}(0, M\frac{(1+\cos^2\phi)}{2})$ and $\mathcal{N}(0, M\frac{(\sin^2\phi)}{2})$ respectively; their covariance is $M(\frac{\cos\phi\sin\phi}{2})$.

2. for $k \neq l$, $E\{u_{ik}u_{il}\} = M(\frac{1+\cos\phi}{2})^2$,
$E\{u_{ik}v_{il}\} = M(\frac{1+\cos\phi}{2})(\frac{\sin\phi}{2})$, and $E\{v_{ik}v_{il}\} = M(\frac{\sin\phi}{2})^2$.

Using the above results, it can be proved that G; and H; are $\mathcal{N}(0, \sigma_G^2)$ and $\mathcal{N}(0, \sigma_H^2)$ and are uncorrelated where

$$\sigma_G^2 = N(N-1)M(\frac{1+\cos\phi}{2})^2 + NM(\frac{1+\cos^2\phi}{2}) \quad \text{and}$$

$$\sigma_H^2 = NM(\frac{\sin^2\phi}{2})$$

Hence
$$q = \int\int_{\mathcal{R}_1} n(x, 0, \sigma_G^2; y, 0, \sigma_H^2; 0)\, dx\, dy$$

By change of variable, we get

$$q = \int\int_{\mathcal{R}_1'} n(u, 0, 1; v, 0, 1; 0)\, du\, dv$$

where $\mathcal{R}_1'$ is bound by lines $v = \pm m'v$ where $m' = (\frac{\sigma_G}{\sigma_H})m$. The angle enclosed by the lines is $2\tan^{-1}[(\frac{\sigma_G}{\sigma_H})m]$ and hence the above integral has the value

$$q = \frac{1\tan^{-1}(m\frac{\sigma_G}{\sigma_G})}{\pi \qquad \sigma_H}$$

q is independent of M and depends only on N and ϕ. The values of q are shown in figure 2. The figure also shows the equivalent of q in the real Hopfield network, which is $\text{Prob}\{F_i > 0\} = 1/2$. From the figure it can be seen that the values of q in the complex model are always less than $1/2$, approaching it when ϕ approaches $180°$.

As already mentioned the calculation of p from q is difficult. Since q for the complex model is less than that of the real Hopfield model, we can expect that the expected number of spurious equilibria is also less.

3 Capacity and content addressability

What is the maximum number of memories that can be stored and retrieved by the model? In this section we answer this question through simulation studies. The capacity of the model depends on the angle ϕ. A network with $N = 125$ was simulated. The performance of the network was studied for $\phi = 90°, 120°, 135°, 150°$, and, 175". For each value of ϕ, the maximum number of patterns retrievable was noted. This results are shown in **figure** 3. It is clear from the figure that as ϕ gets closer to $180°$, the capacity tends to the capacity of the real Hopfield model. We also tested **1000** random vectors for the equilibrium condition in each case none of them were equilibrium points.

The foregoing study concerns the equilibrium properties of the network. Because of asymmetry the general dynamics is difficult to analyze. The model does possess content addressability property. Error correction of up to **15** bits for all memories is possible for a 125-dimensional network, storing **5** patterns with $\phi = 150$'.
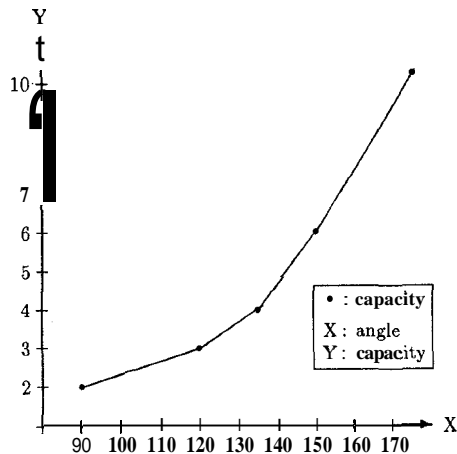

Figure 3: effect of angle on capacity

4 Discussion

We have proposed a model for **CAM** whose performance depends on the parameter ϕ. The performance of the model depends crucially on **d.** We conjecture that the expected number of spurious memories for this class of models varies as $k\,c^N$. We have shown that **c = 1/4** for the random complex asymmetric case; it can be proved that c = 1 for random real asymmetric case; and c = 1.22 for the real symmetric case from a result quoted earlier and the fact that $2^{0.2874N} = 1.2^N$. In the complex model, c depends on ϕ. Since the real model is a special case of the complex model for $\phi = 180°$, and the values of q are less than the real model for $\phi < 180°$ (see figure 2), there should exist a range of values for ϕ for which c < 1. This conjecture remains to be proved.

The capacity of the model is small for small values of ϕ as can be seen form figure **3**. Hence there is a conflict between the selectivity and capacity of the model. The best values of ϕ seem to be in the range 120" to 150" as **N** varies from 20 to 100. This can be see from figure **2** where there is a transition in the values of q from high-low-high in this range.

References

[1] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl Acad. Sci. USA*, vol. **79**, pp. **2554–2558, 1982.**

[2] Y. S. Abu-Mostafa and J.-M. St.Jacques, "Information capacity of the Hopfield model," *IEEE Trans. inform. theory*, vol. **31**, pp. 461–464, **1985.**

[3] P. Baldi, "Neural networks, orientations of the iypercube, and algebraic threshold functions," *IEEE Trans. inform. theory*, vol. **34**, pp. **523–530, 1988.**

[4] R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh, "The capacity of the Hopfield associative memory," *IEEE Trans. inform. theory*, vol. **33**, pp. 461–482, **1987.**

[5] P. Baldi and S. S. Venkatesh, "Number of stable points for spinglasses and neural networks of higher orders," *Phys. Rev. Lett.*, vol. **58**, pp. **913–916, 1987.**

164