A Self-Organizing Neural Network for Locus-addressable Associative Memory

Manoj Raghavan[‡] and Chitra Dorai[t]
Indian Institute of Science, Bangalore - 560012, India.

Abstract

A self-organizing neural network model for locus-addressable associative memory, and binary pattern recognition is presented. The net may be used for either auto-associative or hetero-associative tasks. Locus-addressability is suggested as a possible mechanism for retrieval of memories without any external cues in the form of partial or corrupted exemplar patterns. The architecture, which employs competitive dynamics, embodies a parallel search scheme which updates itself adaptively as the learning progresses. A thresholding mechanism ensures the learning of new exemplars. On saturation of the memory capacity, the net thereafter responds to new patterns by recalling exemplars in its memory that are nearest to the presented input in Hamming distance. The stability-plasticity problem is overcome by 'fast learning' and irreversibility of connection-weight changes. This architecture overcomes the orthogonality and linear independence constraints that limit other models.

1. Introduction

Models for associative memory [1, 23 by and large, presuppose the distributed (holographic) and superposed nature of memories, and recall of memories in these models relies on the presentation of some unique fraction of the stored pattern. It is conceivable that in the brain, regeneration of activity patterns corresponding to different memories is possible by excitation of specific cells or cell groups. The absence of damage tolerance has generally been held to be a major objection to locus-addressable memory. However it is possible to conceive of locus-addressability while preserving damage tolerance, if redundancy of information storage is presumed.

In this paper, a neural network for locus-addressable associative memory, and pattern recognition is proposed with an algorithm for learning and storing binary patterns. We presume the existence of cells whose responses are pattern-specific and demonstrate that a self-organizing network containing such pattern-specific cells can be configured as a content-addresable memory or as a pattern classifier.

2. Design Principles

For the purpose of defining a net for locus-addressable memory, the problem of memory is here phrased as the problem of 'memorising' a neuron's **receptive field** (i.e., neurons from which it receives inputs) and **projection field** (neurons to which it sends its output) at any given time. During the exposure to a given pattern a neuron fixes or 'memorises' its receptive and projection fields, and retains this 'memory' for an extended period of time. Thereafter, the neuron becomes insensitive to the

[‡]Center for theoretical Studies, [t]Dept of Electrical Engineering

firing of neurons outside its receptive field, and also incapable of causing the firing of neurons outside its projection field. Such a neuron becomes a pattern-specific locus which on excitation by even non-specific excitatory stimuli, will regenerate the pattern of activity corresponding to its memory in its projection field. This memorising of fields is achieved here by using specific formulations of synaptic strength alteration rules. Though these do not pretend to be biologically realistic, the synaptic changes depend only on local variables and are activity related.

3. Architecture of the net

The net comprises of four layers: i) an input layer of N units, ii) a thresholding layer of M units, iii) a competitive layer (M units) and iv) an output layer (N units). The behaviour of the neurons in the four layers is controlled by three 'nuclei', I, G, and S which receive their input from one or more layers and feed their output back to one of the layers.

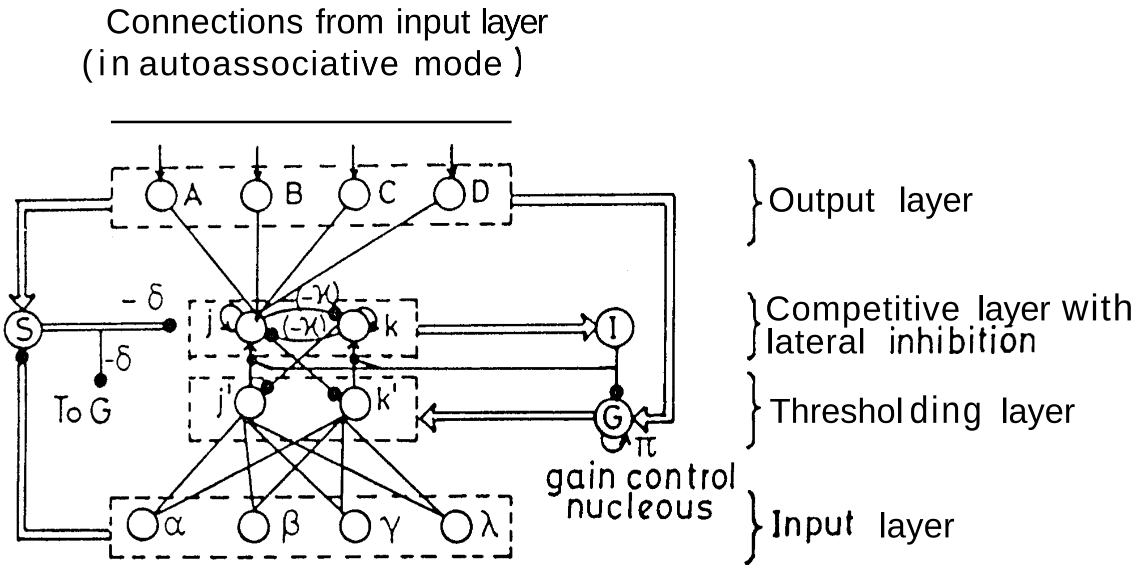

Fig. 1. Net Architecture for 4-bit input patterns & a memory capacity of two patterns. '——>' denotes excitatory connections and —● inhibitory connections.

The input units are denoted by α, β, y,...; units in the thresholding layer i', j', k',...; units in the competitive layer i, j, k, ... and units in the output layer A, B, C etc. The initial thresholds of all units are assigned a value, 0. The weights of all synaptic connections between the competitive and the output layer are initially assigned a value, 0. The weights between the input layer and the thresholding layer, are picked randomly so that $0 < w_{j', \leftarrow \alpha} < 1$ (see figure. 1). There are forward connections of constant value +1 from the cells in the thresholding layer to the corresponding cells in the competitive layer.

If $\xi = [\ \xi_\alpha\ , \xi_\beta\ , \xi_\gamma\ , ..]$, **a** binary vector of N bits **is** presented to the input layer at any given time, then the neurons α, β, y , ... in this layer assume these values as their outputs **so** that $V_\alpha = \xi_\alpha$.

The threshold of a unit j' in the thresholding layer is denoted **as** $T_{j'}$. The net input $\phi_{j'}$ and the output $V_{j'}$ are given by

$$\phi_{j'} \quad = \quad \sum_{\alpha=1}^{N} w_{j'\leftarrow\alpha}\, V_\alpha \quad + \quad \sigma\, V_G \quad - \quad \Omega\ \Big(\ \sum_{\substack{k=1\\ k\neq j}}^{M} V_k\ \Big) \qquad (1)$$

$$V_{j'} \quad - \quad \phi_{j'} \qquad \text{if} \quad \phi_{j'} \ > \ T_{j'}$$

$$= \quad 0 \qquad \text{otherwise} \qquad (2)$$

where $\sigma = w_{j'\leftarrow G}$, is a fixed quantity for all units j' and $\Omega = w_{j'\leftarrow k}$ is a constant of the order of N.

The one to one connections from the units j', in the thresholding layer to the units j in the competitive layer have fixed value of +1.The thresholds for all the cells j, in the competitive layer are zero and j obeys the equation :

$$\phi_j \quad = \quad \theta(1 - \eta\, V_I)\, V_{j'} \quad + \quad V_j \quad - \varkappa\ \Big(\ \sum_{\substack{k=1\\ k\neq j}}^{M} V_k\ \Big) - \delta\, V_S \qquad (3)$$

$$V_j \quad = \quad \phi_j \qquad \text{if } \phi_j \ > 0$$

$$= \quad 0 \qquad \text{otherwise} \qquad (4)$$

where $\theta(x) = 1$ **if** $x > 0$, **and** $\theta(x) = 0$ if $x \leq 0$. The strength of the lateral inhibition between the units in the competitive layer, denoted by $\varkappa$, has a value less than $1/(M-1)$ (see appendix A). The constant δ corresponds to the strength of the connection $w_{j\leftarrow S}$ and is assumed to be greater than N.

In the auto-associative mode, the input layer neurons α, β, y , ... are presumed to send one to one connections of value +1, to corresponding neurons A, B, C .. in the output layer. The net input, and the output of unit A are defined by;

$$\phi_A \quad = \quad \sum_{j=1}^{M} w_{A\leftarrow j}\, V_j \quad + \quad V_\alpha \qquad \& \qquad V_A \ = \ \theta\ (\ \phi_A) \qquad (5)$$

Nucleous I 'gates' input to the Competitive layer by pre-synaptic inhibition. This unit receives input connections from the units j in the competitive layer. All $w_{I\leftarrow j} = +1$. It obeys the equation;

$$\phi_I \quad = \quad \sum_{j=1}^{M} w_{I\leftarrow j}\, V_j \qquad \& \qquad V_I \ = \ \phi_I \quad \text{if } \phi_I \ > 0$$

$$= \ 0 \qquad \text{otherwise} \qquad (6)$$

The synaptic link from the unit I to the unit G is an inhibitory one with strength, η : $\eta > N$. π is a constant weight $w_{G\leftarrow G}$ with a value just greater than 1. The constant weights $w_{G\leftarrow A}$

have a value of ε, $0 < \varepsilon << 1$. The behaviour of G (gain controlling cell) is governed by;

$$\phi_G = \varepsilon \sum_{A=1}^{N} V_A + \pi V_G - \eta V_I - \delta V_S \qquad (7)$$

$$V_G = \phi_G \text{ if } \phi_G > 0$$
$$= 0 \text{ otherwise} \qquad (8)$$

In the absence of any activity in the competitive layer on presentation of an input, (as might happen when saturation of memory occurs and no response is evoked in the thresholding layer by an input) there is no inhibition of G due to I. Since $\pi > 1$, the output of G progressively rises thereby lowering the effective threshold of cells in the thresholding layer. Thus G acts as an automatic gain control unit. This ensures that input patterns always cause recall of exemplars nearest to them in Hamming distance.

Nucleous S renders all neurons which have positive feedback inactive, when the input pattern is removed. This unit receives inputs from all cells in the output layer which are coupled to S via excitatory synapses, $w_{S \leftarrow A}$ and also inhibitory connections, $w_{S \leftarrow \alpha}$ to S from nodes in the input layer. (The weights of the inhibitory synapses are presumed to be greater than the weights of the excitatory synapses, i.e. $w_{S \leftarrow A} < |w_{S \leftarrow \alpha}|$ for all A's and α's). These are constants.

$$\phi_S = \sum_{A=1}^{N} w_{S \leftarrow A} V_A + \sum_{\alpha=1}^{N} w_{S \leftarrow \alpha} V_\alpha \quad \& \quad V_S = \phi_S \text{ if } \phi_S > 0$$
$$= 0 \text{ otherwise} \qquad (9)$$

3.1. Learning Rules

A binary pattern to be stored is clamped on to the input layer till learning is completed. Lateral inhibitory interactions result in the survival of the unit with maximum initial output in the competitive layer. The weights are adapted according to the following rules.

For Receptive field :

$$\Delta w_{j' \leftarrow \alpha} = (1 - w_{j' \leftarrow \alpha}) \, \theta(w_{j' \leftarrow \alpha}) \quad \text{if } \theta(V_{j'}) = 1 \text{ and } V, = 1$$
$$= -(1 + w_{j' \leftarrow \alpha}) \, \theta(1 - w_{j' \leftarrow \alpha}) \quad \text{if } \theta(V_{j'}) \square 1 \text{ and } V, = 0$$
$$= 0 \quad \text{if } \theta(V_{j'}) = 0 \qquad (10)$$

(Recall that $\theta(x) = 1$ if $x > 0$, and $\theta(x) \square 0$ if $x \leq 0$).

For Projection field :

$$\Delta w_{A \leftarrow j} = (1 - w_{A \leftarrow j}) \, \theta(w_{A \leftarrow j} + \mu) \quad \text{if } \theta(V_j) = 1 \text{ and } V_A = 1$$
$$= - (1 + w_{A \leftarrow j}) \, \theta(1 - w_{A \leftarrow j}) \quad \text{if } \theta(V_j) = 1 \text{ and } V_A = 0$$
$$= 0 \quad \text{if } \theta(V_j) = 0 \qquad (11)$$

Here , μ is a small positive quantity : $\mu << 1$.

Learning of thresholds :

We define, $Z = (\psi \phi_{j'} - T_{j'})$ where ψ is a positive constant, < 1. Then $\Delta T_{j'} = Z \, \theta(Z)$.

Since the value of $\phi_{j'}$ at the first presentation of an input pattern is bound to be less than what it would be on a later presentation (due to adaptive weight changes), one would expect the threshold value to stabilize only on a second presentation of the pattern. This can be avoided if updating thresholds is done an iteration after updating the receptive field weights.

4. Self-organization and Pattern Learning

The learning rules presented above ensure 'fast learning' (learn with a single presentation) and the irreversibility of weight changes ensures stability of learnt patterns. The thresholding mechanism of the pattern-specific cells forces the learning of new patterns by uncommitted nodes owing to the fact that the activation of any of the previously commited nodes is prevented unless the presented pattern is close enough to the corresponding exemplar. The refinement in the discrimination of the learning of patterns is determined by the value of the constant ψ .

In the 'fast learning' mode discussed above, it is presumed that the first pattern learnt by each node is taken to be its exemplar. An alternate learning strategy would be to use the same learning rules with a multiplicative factor l/n (where n is some large number) introduced into the learning rules. The use of these fractional weight changes would prevent the weights from reaching irreversible values on a single exposure to a pattern. Thus the net when presented with a series of noisy or corrupted versions of the same pattern, may be expected to form a permanent, stable memory of the features they share.

5. Simulation results

In simulations of this net using 20-bit input patterns along with an equal number of pattern-specific cells, the net was able to learn all twenty presented patterns in a single training cycle. The value of ψ was assumed to be 0.8. The average number of iterations required for convergence during learning was 19. In recall of stored memories the net was found to tolerate upto 40% random noise and the convergence was much faster. The ability of the net to learn and store closely correlated patterns as differnt entities was verified in a simulation in which each of 15 pattern-specific cells learnt its exemplar when the net was presented with 10-bit input patterns. In this case ψ was taken to be 0.9. Random noise of upto 20% was tolerated. In all simulations, during recall, the gain control mechanism was observed to effectively increase the noise tolerance of the net once saturation of memory capacity occurred.

6. Discussion and Conclusion

Unlike perceptrons, and models based on Backpropagation C31 exemplars are learnt by this net in a single training cycle. The **net is** capable of **continuous learning till** saturation of its memory capacity occurs. In its ability to self-organize, and in the competitive nature of learning it is comparable to the adaptive-resonance models of Carpenter and Grossberg [4]. Damage tolerance of memories can be incorporated if we presume the lateral inhibitory interactions to be short-ranged. This would cause more than a single neuron present outside mutual inhibitory range to survive lateral inhibition thereby memorising the same pattern. Though wasteful of neural hardware, this kind of redundancy is biologically conceivable. Another feature of the net is that it may be used for hetero-associative tasks by clamping desired outputs at the output layer while the connection weights are being adapted. In this mode it would be possible for the net to be 'taught' categories, i.e., sets of dissimilar patterns which are to be considered elements of some larger category. The gain control mechanism which becomes effective once saturation of the memory capacity occurs, ensures that the net always responds to a given input irrespective of whether it meets the threshold criteria or not, by recalling the exemplar nearest to the initialization [c.f. 5]. This net has a capacity which scales up satisfactorily with increase in the number of cells in the competitive layer. The network presents a possible way of simultaneously incorporating both **self-organization and stability of** memories.

The locus-addressability of memories suggests interesting potential applications of the net in cognitive modelling. One of these, of current interest to us is the recognition of temporal sequences of inputs and incorporation of context-dependency in the recall of memories, by using a secondary 'association' net.

Appendix A

The requirement for $\varkappa$ to be less than $1/(M-1)$ can be seen from the following; For a unit j with the maximum initial output, the value of ϕ_j is obtained from **(3).** Since $V_S = 0$, (3) reduces to

$$\phi_j = V_j - \varkappa \sum_{\substack{k=1 \\ k\neq j}}^{M} V_k \quad \text{, when I becomes active}$$

If ϕ_j is to remain positive then $V_j > \varkappa \sum_{\substack{k=1 \\ k\neq j}}^{M} V_k$ (12)

Replacing R.H.S by $\varkappa\ (M-1)\ \bar{V}_k$where $\bar{V}_k$ is mean response of unit k we see that $\varkappa < (V_j / \bar{V}_k) * (1 / (M-1))$. Since $V_j > \bar{V}_k$, $\varkappa < 1 / (M-1)$ satisfies the equation (12).

References

1. T. Kohonen, Self-Organization and Associative Memory, Springer Verlag Berlin Heidelberg, 1984.

2. J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," ***Proc.*** *Natl. Acad.* ***Science,*** USA, vol 79, pp. 2554 - 2558, April 1982.

3. D.E. Rumelhart et al, "Learning internal representations by error propagation," Parallel Distributed Processing, vol I, 1986.

4. S. Grossberg, "Competitive learning : From interactive activation to adaptive resonance," ***Cognitive Science,*** 11, PP. 23 - 63, 1987.

5. R.P. Lippmann et al, " A comparison of Hamming and Hopfield neural nets for pattern classification, " MIT Lincoln Laboratory Report, TR - 769.