Published in the Proceedings of the
World Congress on Neural Networks, WCNN'96. July 96

HOW TO MAKE ASSOCIATIVE MEMORIES ATTENTIVE

Javed I. Khan and David Y. Y. Yun
Laboratories of Intelligent and Parallel Systems
Electrical Engineering Department, University of Hawaii
2540 Dole Street, Honolulu, HI-96822, USA
javed@wiliki.eng.hawaii.edu

1.ABSTRACT

This paper investigates parallel and distributed implementation of a class of associative search where during similarity evaluation the scope of the search is dynamically localizable into a sub-set of the pattern elements. This is an important search type with numerous applications ranging from content-based image retrieval, adaptive pattern matching, digital library to vision. Current parallel and distributed neural models based on scalar product rule of synaptic efficacy can not conveniently realize it. In this paper, we identify the prerequisites, analyze the current models and demonstrate a new paradigm of parallel and distributing computing that can realize this search.

2.INTRODUCTION

Traditionally, artificial neural networks (ANN) are known to be applicable in three major application categories, (i) adaptive classification or filter (ii) fast optimization, and (iii) associative memory. However, an intimate look at the success stories of neuro-computing reveals that most of them are confined in the area of adaptive classification or filtering [Carp89, Kulk94]. Hardly any real application flourished which can conveniently take advantage of the associative memory characteristics of neural models.

Since the invention of first artificial neuron by Mcculloch and Pitts, main research emphasis grew in the learning aspect of neuro computing. Increasingly more intricate and complex properties of learning phenomena have been pursued in great depth. Versatility (how arbitrary complex associations can be learned), efficiency (how more patterns can be learned), learnability of causality and temporal relations (Grossberg 1967, Klopf 1987), self-organization (Kohonen 1987, Oja 1982), autonomous unsupervised adaptation (Grossberg 1976, Carpenter & Grossberg, 1987) are just few examples of the intricacies through which research in artificial learning matured [Gros67, Klop87, Koho89, Oja82, Gros76, CGMR92]. Surprisingly, during this period of vigorous emphasis on the learning aspect of ANNs, very few attempts had been made to examine their recollection aspect, other than assuming a very simple model of retrieval. Almost all the proposed learning models since McCulloch and Pitts have been constructed on the assumption of a simple and restricted retrieval scenario, where the sample of the content that is used during query is a close replica of the target. However more complex and versatile retrieval formalism is not only conceivable but also seems to be an integral part of natural associative memories.

In this paper, we take a look into associative computing from the integrated perspective of retrieval and learning. The retrieval capability of our particular interest is the ability of an AAM to dynamically localize match. However, we will also investigate the ability of an AAM to provide a meta-feedback on the quality of match. Dynamic search localization refers to a process where patterns can be retrieved on the basis of similarity within any user given subset of pattern elements. A symmetrical ability is that the external user (human, or any other driver computer) receives a quality feedback to interactively evaluate the patterns regenerated by the memory. These interaction abilities (receive attention information, and provide quality feedback) are important to make associative memories applicable in many computationally daunting problems of today such as object-oriented content based information retrieval, adaptive pattern matching with incomplete pattern, and retrieval with very small cue.

Current artificial associative memories (AAM) based on classical scalar product rule of synaptic efficacy are unable to support either aspects of such retrieval. Recently, an instance of a parallel and distributed computing network with these two critical abilities has been demonstrated by [Khan95]. It is based on a digital adaptation of optical holography and its hyperspheric representation. In this paper we present the result of a broader theoretical investigation that now finds the meta-class characteristics of such a memory, within which other instances of attentive/interactive memories can be invented.

In this paper, we first explain the search types in the context of parallel and distributed associative computing. Then in section 4 we analyze the constraints of current AAMs. In section 5 we demonstrate a new generalized representation and class of cell transfer functions which can overcome these deficiencies. Finally, in section 6 we show the implications of this important search capability which can benefit numerous applications.

3.ASSOCIATIVE MEMORY

Let, $S^{\mu} = [s_1^{\mu}\ s_2^{\mu} \ldots s_N^{\mu}]$ is a stimulus pattern vector and $R^{\mu} = [r_1^{\mu}\ r_2^{\mu} \ldots r_M^{\mu}]$ is any response pattern vector. Here, the superscript refers to index of the pattern vector, and the subscript refers to the particular element in the pattern vector.

Definition (Associative memory): Given a set of stimulus pattern vectors $\mathrm{S} = \{S^{\mu} \mid 1 \le \mu \le P\}$ and a set of response pattern vectors $\mathrm{R} = \{R^{\mu} \mid 1 \le \mu \le P\}$, an associative memory is capable of learning the correspondence between a stimulus member $S^{\mu} \in \mathrm{S}$ and a response member $R^{\mu} \in \mathrm{R}$ in such a way that, given a query pattern S^Q, it can retrieve a pattern $R^R \approx R^T$ such that $R^T \in \mathrm{R}$, and S^Q is closest to $S^T \in \mathrm{S}$ according to a matching criterion D.

An associative memory system (Fig-1) is comprised of (i) a learning algorithm Alearn which converts all the $\{S^{\mu}, R^{\mu}\}$ associations into some internal representation, (ii) a physical storage medium and representation formalism AM to store the associations, (iii) a decoding algorithm Aretrieve to recollect stored information R^R from a given query stimulus S^Q, and (iv) a matching criterion D to measure the closeness of stimulus patterns to the query pattern.

4.ATTENTIVE QUERIES

Concept of pattern distance is central to the search operation of any memory. Equation (1) states a generalized measure of such compositional distance:

$$D(S^a, S^b) = \left[\mathop{M}_{t}^{G} \delta_t \{ dist(s_i^a, s_i^b) \} \right] \text{ where, } \sigma(\cdot) = \left[\mathop{M}_{t}^{G} \delta_t(\cdot) \right]$$

Where, $dist(\cdot)$ is the distance measure function (DMF). It can be any arbitrary function with the constraint that it is monotonic with respect to $|s_i^a - s_i^b|$, and symmetric for all (s_i^a, s_i^b) pairs. M is a set operator with scope G. Generally, a summation is used such that $M = \sum\{\}$. The function $\delta(\cdot)$ is modulator function for the set members. The combined function $\delta(\cdot)$ is the distance composition function (DCF). It can be any function with the constraint that it too is monotonic with $dist(\cdot)$. Finally, the overall function is required to have the property, that $D(S^{\mu}, S^{\mu}) = c$, where c is a constant independent of specific pattern index μ.

Given an sample pattern S^Q an associative memory tries to converge to the closest learned pattern. Let $\Lambda^Q = [\lambda_1^Q\ \lambda_2^Q \ldots \lambda_n^Q]$ be the modulator vector. Consequently, based on the matching criterion of equation (1), we now define a generalized associative memory:

Definition (type-A AAM): Given the modulator vector Λ^Q, a type-A AAM can retrieve response pattern $R^R \cong R^{T_a}$, where its associated stimulus pattern S^{T_a} is close to the query pattern S^Q in the following sense:

$$D(S^Q, S^{T_a}, \Lambda^Q) = \min_{\mu}^{p} \left[\mathop{M}_{t}^{N} \lambda_i^Q dist(s_i^Q, s_i^{\mu}) \right]$$

Modulator vector can be decided at post learning stage, so that the memory can modify the significance or attention level of each pixels dynamically on demand with analog resolution (and will be referred as attention field in the subsequent discussions).

Two other subclasses of this memory with the definitions below are also of interest to us. For the special case, where the modulator vector elements are restricted to binary enumeration, the modulator vector can be substituted by a scope restrictor function $F^Q \subseteq N$ representing a subspace of the total element space N. Finally, a further special case with unary attention of the above two types can be defined where the scope is not restrictable.

Definition (type-B AAM): Given a element subset $F^Q \in N$, a type-B AAM can retrieve response pattern $R^R \cong R^{T_a}$, where its associated stimulus pattern S^{T_a} is close to the query pattern S^Q in the following sense:

$$D\left(S^Q, S^{T_a}, F^Q\right) = \min_{\mu}^{p}\left[\mathop{M}_{t}^{F^Q} \lambda_i^Q dist\left(s_i^Q, s_i^\mu\right)\right]$$

Definition (type-U AAM): A type-U AAM can retrieve response pattern $R^R \cong R^{T_a}$, where its associated stimulus pattern S^{T_a} is close to the query pattern S^Q in the following sense:

$$D\left(S^Q, S^{T_a}\right) = \min_{\mu}^{p}\left[\mathop{M}_{t}^{N} dist\left(s_i^Q, s_i^\mu\right)\right]$$

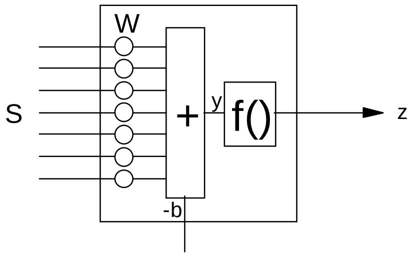

Fig-1 An AM Model

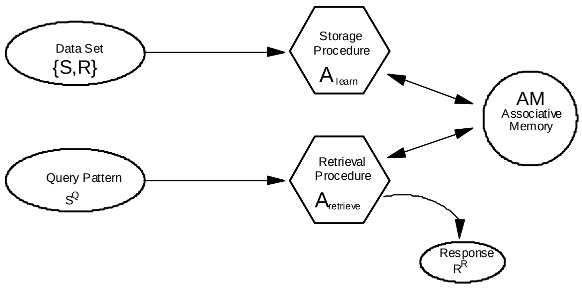

Fig-2 Basic Neuron Cell[1]

Our objective is to investigate the realizability of type-A/B memory within parallel and distributed computing.

5.RETRIEVAL IN CURRENT MODELS

5.1. Current AAM and Artificial Nurode

The variety of architectural configuration and learning techniques that can be interpreted as a distributed and parallel model of artificial associative memory is staggering. However, the cell architecture of Fig-2 and the transfer function of equation set (5) together specify what can be almost unquestionably considered as an essential building blocks of any ANN or AAM.

yi =g(wij,si)=∑wijsi+bi and $z_i = f(y_j)$

[1] An interesting question is that why in the first place, McCulloch and Pitts decided to use a function (transfer function) of the form of a weighted sum as a building block? Possibly, because of its resemblance to the conjunctive normal forms of first order logic.

Here, [si] represents the inputs to this cell and generally they assume values in the range from 0 to 1 or from -1 to 1. This cell has remained virtually unaltered since its invention by McCulloch and Pitts. In this paper, this cell will be referred as MP-neuron, and its transfer function as the scalar product rule of synaptic efficacy (SPRSE). Current AAMs vary among themselves based on (i) the pattern of interconnection network that connects these cells in a network (ii) mode and scenario of learning and/or (iii) the specific type of non-linearity function f(). Within the scope of this paper, we are interested in the underlying function that is optimized by the combined dynamics of the algorithm pair {Alearn:Aretrieve}, or the matching criteria D.

5.2. Retrieval Type

The optimization criteria of existing neural models directly belong to type-U category. Models those use Hebbian class of learning maximize global dot-product of the patterns [Koho89, Gros69, Klop87, Sang89]. On the other hand, the models those use LMS class of learning maximize global mean square error [WiHo60, RuHW86]. There are also other distance measures which have been used in matching criterion (such as likelihood-ratio, entropy, etc.). Hopfield has given a unified perspective and demonstrated that all the neural networks minimize some form of energy function [Hopf82]. The key features to note in all of these functions are (i) the set operator is a summation process $M = \sum \{\cdot\}$, and (ii) the scope G of the set operator is all-element-inclusive and is based on entire element space or G=N, and (iii) the modulator function is only a function of distance $\delta = dist(\cdot)$. These properties of existing neural computation models together makes them a type-U memory.

5.3. Non-Optimality of Conventional AAMs

The reason that conventional AAMs have not succeeded in supporting Type-A or Type-B search lies right at the heart of conventional neurocomputing: the scalar product rule of synaptic efficacy. It is surprising that, despite the invention of so many artificial neural network (ANN) models over the enormously productive fifty years following McCulloch and Pitts, the rule specifying the transformation of signal of a neuron has remained unaltered [Carp94]. We now look into the difficulties of current AAMs. We show:

An associative memory constructed by interconnecting cells with the scalar product rule of synaptic transmission specified by (5) can not realize the retrieval of type-B, or type-A[2].

The demonstration has been constructed in two parts. In the first part, it is shown that a network realizing all element inclusive scope of optimization cannot converge to a correct result with respect to the RCA type-B and type-A search criterion. In the second part, it is shown that the scope of the optimization can not be modified during query for any network which is based on MP-neurons with a SPRSE transfer function.

Part 1 (Problem of All-inclusive Optimization): Let us consider a trained network, which has memorized two patterns $S^1 = [s_1^1 \; s_2^1 \ldots s_n^1]$ and $S^2 = [s_1^2 \; s_2^2 \ldots s_n^2]$. Let us divide the set of total element space N into two arbitrary subsets A and B, such that $A \cup B = N$. Without loss of generality let us also assume:

$$0 << \left[\sum_i^A dist\left(s_i^1, s_i^2\right) \right] < \left[\sum_i^B dist\left(s_i^1, s_i^2\right) \right]$$

Now, let us consider a query stimulus, $S^Q = [s_1^Q \; s_2^Q \ldots s_n^Q]$ carefully constructed in such a way that:

$$\begin{aligned} s_i^Q &= s_i^1 \text{ when } i \in A \\ &= s_i^2 \text{ when } i \in B \end{aligned}$$

Let us also consider an attention distribution vector, $\Lambda^Q = [\lambda_1^Q \; \lambda_2^Q \ldots \lambda_n^Q]$ where:

$$\begin{aligned} \lambda_t^Q &= 0.0 \; when \;\; i \in A \\ &= 1.0 \; when \;\; i \in B \end{aligned}$$

[2] Procedural approaches for type-A/B search is also prohibitively expensive. Please see [Khan95] for detail complexity comparisons.

Now, first investigate the optimum result expected from the query of type -A. From the initial state S^Q, the distance measure according to a search of type-A, between S^Q and S^1 is:

$$D\left(S^Q, S^1, \Lambda^Q\right) = \left[\sum_i^N \lambda_i^Q dist\left(s_i^Q, s_i^1\right)\right] = \left[\sum_i^B dist\left(s_i^Q, s_i^1\right)\right] > 0$$

And, the distance from the second stimulus is:

$$D\left(S^Q, S^2, \Lambda^Q\right) = \left[\sum_i^N \lambda_i^Q dist\left(s_i^Q, s_i^2\right)\right] = \left[\sum_i^B dist\left(s_i^Q, s_i^2\right)\right] = 0$$

Thus, from (7a) and 7b):

$$D\left(S^Q, S^1, \Lambda^Q\right) > D\left(S^Q, S^2, \Lambda^Q\right)$$

Which implies the expected result is $R^R \cong R^2$.
Now let us see the actual output of a cell with above definition. The distance between the patterns according to estimate A:

$$\begin{aligned} D(S^Q, S^1) &= \sum_i^N dist(s_i^Q, s_i^1) \\ &= \sum_i^A dist(s_i^Q, s_i^1) + \sum_i^B dist(s_i^Q, s_i^1) \\ &= 0 + \sum_i^B dist(s_i^Q, s_i^1) \end{aligned}$$

Similarly, the distance measure between S^Q and S^2 is given by:

$$\begin{aligned} D(S^Q S^2) &= \sum_i^N dist(s_i^Q, s_i^2) \\ &= \sum_i^A dist(s_i^Q, s_i^2) + \sum_i^B dist(s_i^Q, s_i^2) \\ &= \sum_i^A dist(s_i^Q, s_i^2) + 0 \end{aligned}$$

An optimally trained network of such cells will converge to a pattern which is at minimum distance from the query. A learning based on LMS rule or its variant will converge in a least mean square error solution, on the other hand a network, with Hebbian learning or its variant will converge to maximum dot product solution. Due to inequality (6), in both cases:

$$D\left(S^Q, S^2\right) > D\left(S^Q, S^1\right)$$

Thus, from (8a) and (8b), the produced result will be $R^R \cong R^1$. Which is a clear contradiction to the expected result from search of type-B or type-A. (proved)

Part 2 (Scope Inflexibility of SPRSE neuron): Let us consider, the role of any ith neuron in the network. Let us also consider that to reconstruct the expected pattern R^2, its corresponding ideal output is $Z_i^{ideal} = f(y_i^{ideal})$. The corresponding ideal input vector is $S^{ideal} = [s_1 \ s_2 \ldots s_N]$. Let, the learned weight vector is $W^{ideal} = [w_1 \ w_2 \ldots w_N]$. Therefore, the ideal weighted summation output of the cell is:

$$y_i^{ideal} = \sum_j^N w_{ij} \cdot s_i + b_i$$

Now, if the attention vector Λ^Q is imposed on it as a scope constraint, then the modified and constrained output of the cell becomes:

$$y_i^{const} = \sum_j^B w_{ij} \cdot s_i + b_i$$

Thus, the internal error is:

$$y_i^{error} = y_i^{ideal} - y_i^{const} = \sum_j^A w_{ij} \cdot s_i$$

Individual terms in this summation are in the order of $O(w \cdot s)$. In addition, if:

$$B << N, \text{or} \sum_i^B \lambda_i << N$$

Then the overall summation itself will be in the order of $y^{error} \cong O(y^{ideal})$. For large $N \cong A >> 1$ the sum will behave like a random walk and the expected value of its growth will be of the order of $|ws| \cdot \sqrt{N}$. The exact external error (at z) will depend on the specific type of the activation function $f()$. For any non-linearity with unimodal first derivative function (which includes all sigmoidal and step non-linearity used by conventional AAMs), small internal errors will be corrected but larger errors will be magnified. Thus, the actual output of this cell will also be numerically off-balanced from the ideal output in the order of:

$$z^{error} = g(z_i^{ideal}) - g(z_i^{const}) \approx O(z_i^{ideal})$$

The above analysis is true for any cell in a network. For $z^{error} \cong O(z^{ideal})$, a network of non-linear SPRSE neurons will run into avalanche magnification of error. Therefore, collectively a network made of SPRSE rule will fail to converge when the scope is altered.

Synaptic Efficacy and RCA

The formal analysis of the previous section can also help identifying the underlying causes of the attention deficiency and provide important insight to its possible solutions. We briefly summarize the key aspects here:

(i) What is generally referred to as the 'robustness' of a ANN originates from the effect of activation non-linearity. However, the same non-linearity that helps in correcting error may also catastrophically amplify error.

(ii) The size of the error depends on the statistical balance between the 'correct' versus 'incorrect' components of error. For correct recall, the cue signal strength in the query-pattern must be statistically dominant over the strength of rest of the pattern elements.

(iii) The exact weight of a particular erroneous element is decided by fixed vector W. Vector W is pre-decided during learning and can not be modified dynamically at query. This eliminates the possibility of multiplicative modification like making of synaptic inputs in proportion to their attention.

(iv) What is generally known as robustness of ANN, is more specifically its robustness against noisy input. MP-neurons and the collective network built upon them, does not have direct mechanism to be robust against missing elements.

6.SYNAPTIC TRANSMISSION RULES FOR TYPE-A MEMORY

The above discussion reveals the difficulties of the MP-neurons and the networks built upon them, to cope with the query-modulated attention over the stimulus pattern space. It also provides strong indication that simple architectural modification cannot solve this non-optimality of the cell. In fact, there is no representation framework which can accept meta-quantity attention or return meta-quantity confidence feedback. It only represents and processes the measurement component of information.

6.1. Representation for Interaction

The first step towards the solution is embedded in the representation. One of the principal requirement is to convey to the network the notion of attention. Consider the special case of "don't care" (equivalent to no attention). Can we have a representation for "don't care"? It is clear from the previous discussion that if s's are variable enumerated by real numbers in the finite real interval I=[1,0], the use of any real value d in this interval to convey the notion of "don't-care" to the cell can not purport the intended action (true even for d=0, which, may appear as a simple

solution to this deceptively hard problem at the first look!), but merely introduces uneven bias towards two other attractors in proportion to their distance from d.
**

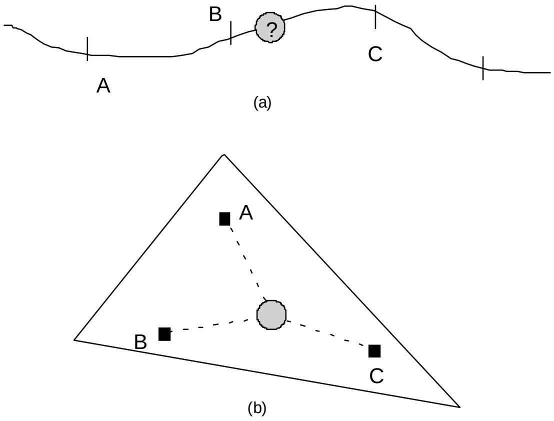

Fig-4 Representation

In any linear state-space it is not possible to obtain a point which is equidistant from all possible enumerations of an analog measurement. Any enumeration of 'dont-care' (denoted by the circle in Fig-4(a)) on a real line will always induce undue bias towards two of the enumerations than all others (such as towards B and C, than A).
A solution to this problem is to place the enumerations on a plane (Fig-4(b)). For example a state-space allowing the enumerations to be laid out on the vertices of a equilateral triangle will allow us to obtain a point equidistant from three points, thus, allowing the construction of a tri-state memory with attention. Similarly, a square state-space will allow the construction of a quad-state type-A/B memory. A circular state-space will allow construction of a analog type-A/B memory. Where the distance from the "point of equidistant" can be used to represent the attention/confidence. In its most generalized form sphere can be used. Thus, a representation suitable for meta-knowledge transformation should of the following form, where H is a n-dimensional (n>1) state space.

$$s_i \Rightarrow H(\theta_i, \beta_i)$$

Here θ represents a state enumerating the measurement component of information and β represents the meta-knowledge (attention for input/ memory confidence for output) component of pattern elements. Importantly, this same representation scheme can also be used by the memory to convey the quality feedback. This is because, this feedback in just a symmetrical quantity to attention.

6.2. Analog Summation Process

As evident from the previous analysis that the next major difficulty arises from the algebraic summation process itself* in SPRSE. An analog summation process requires all the components to be present for its reconstruction. Absence of any element in the summation set can critically hurt the outcome. This is particularly true for finite size of the set. What types of set operator functions function can be used which can have robustness in this sense?
An averaging process is certainly one of the candidate with the above quality. However, there is one subtle requirement. The overall cell transfer function must be non-linear. For a system trying to be robust in the face of mission elements non-linearity should be local to inputs rather than global. Let, all the quantities are now represented in multidimensional state space H. So we will use phase notations to denote the enumerations of cell output and input (compared to equation (5) $\theta_i \approx z_i$ and $\theta_i \approx s_i$). Also let the learned weight:

$$\vec{w}_{ij} = \left\| w_{ij} \right\| \cdot e^{-iw_{ij}}$$

Where the magnitude represents the distance of the learned state from the center (point of "dont-care"), and the phase represents the learned enumeration. The following generalized class of "averaging" transfer function can be used to this purpose:

$$\phi_i = F^{-1}\left[\frac{1}{c}\sum_j^n \left\| w_{ij} \right\| \lambda_j F\left(dist\left(\theta_i, w_{ij}\right)\right)\right] \text{ where, } c = \sum_j^n \left\| w_{ij} \right\| \cdot \lambda_j$$

Here, $F()$ and its inverse $F^{-1}()$ determine the nature of the discriminating hyperplane that maps stimulus patterns on to the response classes, and c is the normalization operator which is in some sense a sum of the dynamic strengths of the inputs present. In each cell, the magnitude (β_i component of the output should be computed in proportion of the evaluated distance. The multiplicative modification of the contribution of individual elements through modulator elements can be used to dynamically control the contribution of each element. Various instances of learning algorithms can now be designed to learn w_{ij} in H within this framework of cell dynamics.

7.CONCLUSIONS

In this paper we have analyzed the realizability of a memory which can perform search with localized attention within the paradigm of parallel and distributed associative computing. This type of memory is applicable in numerous pattern matching and memory retrieval problems. Some of its important implications are illustrated below.

Object/Feature based Search: In any general situation when we describe similarity, in fact the similarity is based on some assumed features or objects of the pattern space rather than global pixel-to-pixel similarity. Any search that can accommodation such underlying assumptions of object based similarity would require localization.

Dynamic Attention: One of the most important aspect of localization based retrieval that we are concerned with is the dynamic specifiability of the field of localization. If a specific distribution of attention is given during encoding at pre-learning stage, a conventional AAM might in some situations (if it is also reflected in the statistics of the training examples) is able to hard-encode it in the learned synaptic weights. However, once the learning is over, the distribution of attention can not be recast during query. For a given learning, it acts as a deterministic machine where each initial state flows into a pre-determined single attractor. Conventional AAMs have no mechanism to accommodate post-learning change in the distribution of attention*.

Managing Incomplete Information: Dynamically localizable search is also critically important in the processing of incomplete information. Notably, this type of imperfection is quite different from the case of noisy information. In former case, some extra information is available about the location of the missing elements. A pattern matching machine which can not localize its search, can not take advantage of this extra information and consequently converges to a solution of lesser optimality for the available amount of knowledge[3].

Statistical Strength of Effective Cue: A serious consequence of the inability of localize search is the inability to work with a small cue. For almost all NNs, as the number of error bits approach approximately just 40%, the probability of correct recall vanishes sharply irrespective of the sophistication of the learning algorithms. Experiments published by many researchers contains the fingerprint of such pure statistical nature of neural network convergence* [TaJo90, Hopf82, MiFa90, KuWo91] (although, it apparently always managed to escape pursuation). Apparently, there lies a fundamental statistical dominance barrier close to 50%. Clearly, this is a profound limitation (from both the biological and the practicability rationale of AAMs) for any effective memory (like the biological memories). A memory with the ability of search localization can potentially overcome this limitation by obtaining match only with local dominance within the field of attention. A locally dominant cue can be globally small.

An instance of this class of memory has been recently demonstrated by [Khan95], and applied for content-based image retrieval. Currently, we are experimenting two other attentive learning models based on principal component analysis [KhYu94,Sang89], and self-organizing feature map [Koho89]. Diverse models of attentive memories are potentially realizable (to suit specific applications) within the generalized representation and synaptic efficacy function family demonstrated in this paper with appropriate transformation of many current learning models.

[3] Hopfield in his famous 1982 paper wrote "..[patterns in] memories too close to each other are confused and tend to merge... For N=100 [number of elements in a pattern], a pair of random memories should be separated by [at least] 50ñ5 Hamming units". Which actually indicates to the fact that for correct operation, the query pattern should have at least 50% similarity to the target pattern.

8.BIBLIOGRAPHY

[Carp89] Carpenter, G. A., "Neural Network Models for Pattern Recognition and Associative Memory", Neural Networks, v.2, 1989.

[CGMR92] Carpenter G. A., S. Grossberg, N. Markuzon, J. H. Reynolds, & D. B. Rosen, "Attentive Supervised learning and Recognition by Adaptive Resonance Systems", Neural Networks for Vision and Image Processing, Ed. G. A. Carpenter, S. Grossberg, MIT Press, 1992, pp364-383.

[Gros67] Grossberg, S., "Nonlinear Difference-Differential Equations in Prediction and Learning Theory", Proc. Of National Academy of Science, v.58, n.4, October 1967, pp1329-1334.

[Gros69] Grossberg, S., "Embedding Fields: A Theory of Learning with Physiological Implications", J. of Mathematical Psychology, v.6, 1969, pp209-239.

[Gross76] Grossberg, S., "On the Development of Feature Detectors in the Visual Cortex with Applications to Learning and Reaction-Diffusion Systems", Biological Cybernetics, v.21, n.3, 1976, pp145-159.

[Hopf82] Hopfield, J. J., "Neural Networks and Systems with Emergent Collective Computational Abilities", Proc. of National Academy of Science, USA, v.79, April 1982, pp2554-2558.

[Khan95] Khan, J. I., "Attention Modulated Associative Computing and Content Associative Search in Images", Ph.D. Dissertation, Department of Electrical Enginnering, University of Hawaii, July, 1995.

[KhYu94] Khan J. I.,& D. Yun, "Feature Based Contraction of Sparse Holographic Associative Memory", Proceedings of World Congress on Neural Networks, WCNN'94, v.4, San Diego, June, 1994, pp26.

[Klop87] Klopf, A. H., "Drive-Reinforcement Learning: A Real Time Learning Mechanism for Unsupervised Learning", Proc. of 1st IEEE Conf. on Neural Networks, Vol.II, N.J., 1987, pp441-445.

[Koho89] Kohonen, T., Self-Organization and Associative Memory, 3rd Ed., Springler Verlag, Berlin, 1989.

[Kulk94] Kulkarni, A. D., Artificial Neural Networks for Image Understanding, Van Nostrand Reinhold, New York, 1994, pp15.

[KuWo91] Kumar, B. V. K., P. H. Wong, "Optical Associative Memories", Artificial Neural Networks and Statistical Pattern Recognition, I. K. Sethi and A.K. Jain (Eds.), Elsevier Science Publisheres, 1991, pp219-241.

[MiFa90] Michel, A. N., J.A. Farrel, "Associative memories via Artificial Neural Networks", IEEE Control Systems,v.10,no.3, April 1990, pp6-17.

[Oja82] Oja, E., "A Simplified Neuron Model as a principal Component Analyzer", Journal of Mathematical Biology, v.15, 1982, pp267-273.

[RuHW86] Rumelhart, D.E., G.E.Hinton, R.J.Williams, "Learning Representations by Backpropagation Errors", Nature, 323, 1986, pp533-536.

[Sang89] Sanger, T. D., "Optimul Unsupervised Learning in a Single layer Linear Feedforward Network", Neural Networks, v.2,1989, pp459-473.

[Tajo90] Tai, Heng-Ming, T. L., Jong, "Information Storage in High-order Neural Networks with Unequal Neural Activity", J. of Franklin Institute, v.327, n.1, 1990, pp16-32.

[WiHo60] Widrow, B., M.E. Hoff, "Adaptive Switching Circuits", IRE WESCON Convention Record, part 4, 1960, pp96-104.