Self-Organization in Artificial Intelligence and the Brain

Ananth Ranganathan, Zsolt Kira
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332
{ananth, zkira}@cc.gatech.edu

***Abstract*— Self-organization is one of the few theories that can explain significant aspects of developmental neuroscience. Within the brain itself, various spatially organized regions, or maps, exist that emerge dynamically. Theories and models that use self-organization have been successful at explaining such phenomena, and while these are not conclusive proof, they provide strong evidence in favor of self-organized mechanisms in the brain. Artificial Neural Networks have been developed that make use of these models to produce pattern recognition and classification mechanisms that have been used in widely diverse fields. This paper describes some of the models used to explain the emergence of various patterns and maps in the brain and their counterparts in the Neural Network domain. Widely used Neural Network algorithms include the Self-Organized Map and Adaptive Resonance Theory, that are discussed herein.**

I. Introduction

There are many reasons for believing that self organization is a fundamental mechanism used in the brain. The brain consists of approximately 10^{10} neurons, each of which can have up to 10^4 connections. Such a complex system cannot be encoded using genetic information alone. Furthermore, the brain is not a static object; it changes through its interaction with the environment, especially during development but even in adulthood. There is also some neuroscientific evidence for self organized processes, such as the success of models using Hebbian learning and the various patterns and maps that emerge dynamically in the brain. For example, various models using self-organization have successfully reproduced patterns found in the visual cortex.

Despite this, it is not known whether other methods of organization are also used. For instance, neurons can release neurotransmitters into the extracellular fluid which diffuse and can be received by a group of neurons [Kelso 1995]. Nonetheless, any form of central control cannot account for the incredible amount of processing that goes on in the brain. It would be impossible to keep track of all neural connections and since the plasticity of the brain has been well established, it cannot be formed from any sort of template or recipe.

After these findings of self-organization in the brain were discovered, they were embraced by the artificial intelligence community as something that could provide clues as to what intelligence really is. The brain is an obvious model for the construction of artificial intelligent agents. Early research into the brain revealed a structure comprising a complicated inter-communicating network of billions of neurons, with a relatively simple structure for the neuron itself [McCulloch and Pitts 1943]. The field of connectionism was fostered upon the belief that global intelligent behavior, such as memory, pattern recognition etc., resulted from local interactions among a large number of simple processing units working independently. While the simple neuron model has long been overthrown in the neuroscience community, the success of neural networks in diverse application areas has ensured continued interest in them among researchers in artificial intelligence.

The field of artificial neural networks can be said to have begun with the seminal paper by McCulloch and Pitts [McCulloch and Pitts 1943], that described how neurons in the brain might work, and included a simple neural network model based on electrical circuits. In 1959, Bernard Widrow and Marcian Hoff of Stanford developed models called ”ADALINE” and ”MADALINE” [Widrow and Hoff 1960], which stand for ADAptive LINear Element and Multiple ADAptive LINear Element respectively. Multi-layered feed-forward networks, back-propagation networks, and hybrid networks, which were developed subsequently, use ADALINE-like components as their basic computations units.

Though the original work by McCulloch and Pitts was biologically motivated, it is generally accepted that perceptrons, back-propagation and many other techniques are not biologically plausible [Kaplan 1991]. In parallel with the above work, networks based on self-organization, that used biologically-plausible techniques such as competitive learning, were also being developed. The pioneering work in this regard was by von der Malsburg and Grossberg [von der Malsburg 1973][Grossberg 1976a]. Fukushima built biologically inspired networks, called Cognitron and NeoCognitron, that explored related work [Fukushima 1975][Fukushima 1980]. In the 1970s, Grossberg developed his Adaptive Resonance Theory (ART) [Grossberg 1976b], which contained a number of novel hypotheses about the underlying principles governing biological neural systems. These ideas formed the basis for the three classes of ART architectures developed by Carpenter and Grossberg. These are self-organizing neural implementations of pattern clustering algorithms and are discussed in greater detail below. Competitive learning, with lateral feedback, is also the basis for Kohonen's Self-Organizing Feature Map (SOFM) and Learning Vector Quantization (LVQ) algorithms. The SOFM is unique in effectively creating a spatially organized internal representation of various features of the input signal.

The rest of the paper is organized as follows. After a brief overview of neurons in section 2, we focus on two main aspects of self-organization in natural and artificial intelligent systems. The first of these, in section 3, is self-organized learning, which contains descriptions of Hebbian learning, as applicable to the brain, and Competitive learning algorithms, that are applicable to artificial neural networks. The description of Hebbian learning includes simple Hebbian learning, as well as extensions that make the learning competitive. The Adaptive Resonance Theory, which uses competitive networks to solve the stability-plasticity dilemma in artificial neural networks, is also presented. The other aspect of self-organization, discussed in section 4, relates to models of the various topographic maps found in the brain, including extensions to the models created by researchers in artificial intelligence that allow networks to do useful things, such as spatially ordering data spaces. The paper ends with an analysis of the various mechanisms involving self-organization in the brain and neural networks and how they relate to other types of self-organization.

II. Neurons

At the cellular level, the brain and nervous system are composed of a vast network of interconnected cells called neurons. Neurons can be of many types and shapes, but ultimately they function in a similar manner. If the brain is self-organized, neurons would be the individual elements that interact in order to form a global pattern. As can be seen from Figure 1, neurons contain long and short extensions called axons and dendrites, respectively. The neurons are connected to each other through these extensions. Dendrites carry electric potentials towards the cell and axons carry them away from the cell. The dendrite of one cell is connected to the axon of another, with a small gap in between called the synaptic gap or synapse. In order to transmit information from one cell to another the cell transmits an electric signals that travels down the axon and causes the release of neurotransmitters that travel through the synapse to the other cell. Note that this type of interaction is completely local. The connections among neurons are not static, however. Connections are strengthened and weakened constantly, and this type of interaction forms the basis of learning. Although they are the basic units that make up the brain, neurons are far from simple by themselves. There is an abundance of models that try to predict and explain the temporal and electrical characteristics of neurons (for example, [Ruf and Schmitt 1998] [Gerstner and Kistler 2002]).

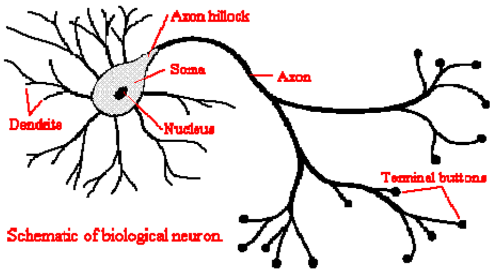


Fig. 1. A neuron

III. Self-Organized Learning

How do we manage to learn and remember complex sequences of events that consist of multiple types of sensory information? This integration of sensory information during learning is all the more astonishing when we consider that two neurons in, say, the visual and auditory cortices of the brain, are separated by billions of intervening neurons and share no common synapses that can lead to easy integration of information. Another important aspect of learning is the ability to distinguish important information from unimportant information. This section tries to provide insight into learning in self-organized systems such as the brain. Hebbian learning, and its counterpart in artificial neural networks - Competitive learning, are presented. Adaptive Resonance Theory, which uses competitive learning along with some other rules to overcome the stability-plasticity problem, is also discussed.

Hebbian Learning

The fundamental assumption used in most models of the brain is the existence of Hebbian synaptic plasticity, or Hebbian learning. In 1949, a Canadian researcher named Donald Hebb proposed a mechanism by which neurons change the strength of their connections to other neurons. The rule is as follows [Brown and Chattarji 1995]:

> When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.

The original postulate has been modified and clarified since its inception in order to add some key characteristics. The first characteristic is that this mechanism is local: the neurons respond to local information through their connection to neighboring cells. This does not rule out global control signals that may be used to control Hebbian plasticity in a group of cells. There is some evidence that neuromodulators may act in this role. A second important characteristic is that the interaction requires activity on both sides of the synapse. This results in neurons that act on correlated input to reinforce one another. The final feature of this model is that the correct timing of the pre-synaptic and post-synaptic activity (activity of cell A and cell B, respectively) is essential in determining how the connections are modified.

Hebbian learning is one of the most important concepts used for unsupervised learning in Neural Networks. One use of networks using this rule is in creating an associative memory [Brown and Chattarji 1995]. An associative memory is a system that can "recall" mappings between specific inputs and specific outputs. Also, it has been shown that the standard implementation is optimal in finding correlations under the assumption of Gaussian noise. Suppose that there is a network of nodes N_i, each connected to other nodes and the strength of the connection between node i and node j is W_{ij}. The simple Hebbian rule can be written as:

$$y_i = \sum_j w_{ij} x_j \tag{1}$$

$$\Delta w_{ij} = \eta x_j y_i \tag{2}$$

Equation 1 simply states that y_i, the output from neuron i, is equal to x_j, the input from neuron j (for all neurons j connected to i), times w_{ij}, the weight of the connection between neuron i and neuron j. Equation 2 is the Hebbian learning rule, which states that the change in weight between neuron i and j is influenced by the learning rate η, the input received from neuron j (pre-synaptic), and the output of neuron i (post-synaptic). The learning rate is usually a small number that can be decreased through time. Notice that this means that if the two neurons fire simultaneously, then the weight of their connection will increase proportionally to the strength of firing. If Equation (1) is substituted into (2) then the result is

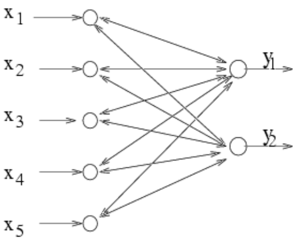


Fig. 2. The negative feedback network. Activation transfer is fed forward and summed and returned as inhibition. Adaption is performed by simple Hebbian learning [Fyfe 1997].

$$\Delta w_{ij} = \eta x_j \sum_k w_{ik} x_k = \eta \sum_k w_{ik} x_k x_j \tag{3}$$

It can be shown using this last equation that the network using this rule is able to find correlations in the data set [Fyfe 1997]. This simple rule is not enough, however, because it is unstable; repeated use can increase the weights of the connections without bounds, and the performance will degrade since all the neurons will be saturated to their maximum values. This is in part due to the positive feedback in the system: Larger weights will results in a larger output, which will result in a larger increase of weights. It is also biologically implausible since there is a limit on the number and efficiency of synapses per neuron.

As a result, Oja and others [Oja 1982] [Oja 1989] have come up with rules that have a decay term, which can be implemented using negative feedback. The resulting network can be seen in Figure 2. Equation 1 remains the same, but the firing from one time cycle is fed back to the next as inhibition to give

$$x_j(t+1) \leftarrow x_j(t) - \sum_{k=1}^{M} w_{kj} y_k \tag{4}$$

If the same substitutions as before are made, the resulting equation is

$$\Delta w_{ij} = \eta_t y_i x_j(t+1) = \eta_t y_i \left(x_j(t) - \sum_{l=1}^{M} w_{lj} y_l \right) \tag{5}$$

Here, $x_j(t)$ and $x_j(t+1)$ differentiates between activation at times t and $t+1$. With these specific changes, the weights will converge and it has been shown that the networks ends up doing primary component analysis (PCA) [Fyfe 1997]. PCA finds the best linear compression of data by finding the linear basis of a data set that minimizes the mean squared error between the compressed and uncompressed data.

Another method to overcome the problem of infinitely increasing connection weights is to renormalize the total synaptic weights of all the inputs, so that the total input weight is a constant. This results in competition, since an increase in weight from one neuron results in a decrease in the weights of connections to other neuron. There have been many variations of Hebbian learning that result in very different patterns or functions resulting.

Although Hebb could not verify his theory, some evidence for Hebb's rule was later found in the hippocampus in the form of Long Term Potentiation (LTP). In studies of animal brains, it was found that certain stimulation of axons in the hippocampus lead to an increase in the synaptic strength as measured by the post-synaptic response. This phenomenon, named Long-Term Potentiation, was found to last anywhere between a few hours and a few days. LTP is an extension of Hebb's rule and states that if a weak and a strong input act on a cell at the same time, the weak synapse becomes stronger [Gazzaniga et. al. 2002]. In order for the synapse to be strengthened, the activity must occur at the same time. Also, if a second weak input exists but is not active while the strong synapse is active (unlike the first weak input), it will not be strengthened. There has been a great deal of research that has attempted to explain this phenomenon in terms of ion flows between the neurons, although that is beyond the scope of this paper.

Competitive Learning

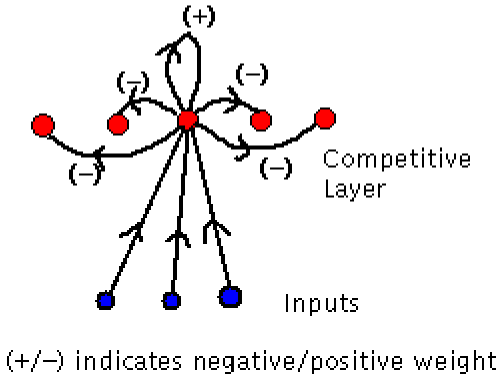


Fig. 3. A single node in a network illustrating competitive dynamics

Competitive learning is an algorithmic implementation of Hebbian learning in artificial neural networks. With reference to neural networks, it refers to a family of algorithms that use some sort of competition between lateral nodes during learning. Normalization of the sum of inputs to a neuron, as done in Hebbian learning is one simple form of that. This section will discuss more complicated forms that have some useful properties. In a typical competitive network architecture, nodes in each layer are connected to a layer above as in traditional neural networks, but in addition, there are lateral connections between nodes in the same layer which cause the competition. Competitive learning includes a wide variety of algorithms performing different tasks such as encoding, clustering, and classification.

Competitive learning is also referred to as self-organized or unsupervised learning. This presents it in contrast to supervised learning (for eg., as in back-propagation), where the network receives feedback in the form of target outputs or environmental evaluations for each input. In unsupervised learning, the learner decides which aspects of the input signal structure to capture in the output. In essence, the learner builds a probabilistic model of input and uses this to generate a recognition distribution, given a specific instance of the input. Feedback in competitive networks occurs at two levels. Firstly, the competitive network utilizes feedback, in the form of lateral inhibitory connections and self-excitation, to pick the winner of the competition. Subsequently, the winner's weight vector is changed to minimize the error between the input and the weights. Thus, the error function is itself a feedback signal that informs the network of the direction in which change is required.

In recent years, competitive learning has received considerable attention due to its demonstrated applicability and biological plausibility. Competitive learning algorithms employ competition through lateral connections between nodes in the same layer. In early models, the competition, called hard competition, led to final activity of a single node, the strongest one to start with [von der Malsburg 1973]. More recent models do not drive down all but one of the nodes to zero, even though only one emerges the winner from the competition [Nowlan 1991]. This form of competition, called soft competition, has many advantages over hard competitive methods.

A general problem occurring with hard competitive learning is the possible existence of "dead units". These are units which, perhaps due to inappropriate initialization, are never winners for any input signal and, therefore, keep their position indefinitely. Those units do contribute to the network functioning and must be considered harmful since they are unused network resources. Another problem with hard competitive learning is that different random initializations may lead to widely differing results. The purely local adaptations may not be able to get the system out of the local minimum where it was started. Soft competitive methods overcome these problems and, hence, are preferred over hard competitive methods for most purposes. However, the comparative simplicity of hard competitive methods makes them good instruments for illustrating competitive learning.

The biological plausibility of competitive learning can be inferred from the fact that the nodes in competitive network develop into feature-sensitive detectors. Feature-sensitive cells are also known to be common in the brain, though the extreme views suggested by proponents of "Grandmother" cells have been discarded. Neural modelers have been able to suggest processes through which such cells can emerge from simplified membrane equations of model neurons [Nass and Cooper 1975][Perez et. al. 1975]. There is evidence that inhibition plays a similar role in various brain functions, such as the creation of orientation columns in the visual cortex [Ramoa et. al. 1986] or sharpening receptive fields in the sensory cortex [von der Malsburg 1973][Merzenich et. al. 1988].

Network Evolution Through Competitive Dynamics: An Illustration

A competitive learning network comprises the feed-forward excitatory network and the lateral inhibitory network. The feed-forward network usually implements an excitatory Hebbian learning rule. The lateral competitive network is inhibitory in nature. The network serves the important role of selecting the winner, often via a competitive learning process highlighting the "winner-take-all" schema. In a winner-take-all circuit, the output unit receiving the largest input is assigned a full value(e.g.,1), whereas all other units are suppressed to a 0 value.

Consider a layer or group of units as shown in Figure 3. Each cell receives the same set of inputs from an input layer. There are intra-layer or lateral connections such that each node is connected to itself via an excitatory (positive) weight and inhibits all other nodes in the layer with negative weights. Now suppose a vector **x** is presented at the input. Each unit computes a weighted sum **s** of the inputs provided by this vector. That is

$$\mathbf{s} = \sum_i w_i x_i \qquad (6)$$

Some node k, say, will have a value of **s** larger than any other in the layer. It is now claimed that, if the node activation a, is allowed to evolve by making use of the lateral connections, then node k will develop a maximal value while the others get reduced. The time evolution of the node is usually governed by an equation which determines the rate of change of the activation. This must include the input from the lateral connections as well as the 'external' input given by **s**. Thus if l is the weighted sum of inputs from the lateral connections

$$\frac{da}{dt} = \beta_s s + \beta_I l - \gamma a \qquad (7)$$

where β_s and γ are positive constants, β_I is negative and the γ term denotes activation decay. It can be observed that the node k with greatest excitation **s** from the input has its activation increased directly by this and indirectly via the self-excitatory connection. This results in inhibition of the neighboring nodes, whose inhibition of k in turn, is then further reduced. This process is continued until stability is achieved. There is therefore a 'competition' for activation across the layer and the network is said to evolve via competitive dynamics.

The Learning Technique

During the learning phase, the network is presented with a training set of inputs. In general, each input is a vector. The goal for the network is to learn a weight vector configuration that minimizes the total error between the weight vectors and the training set. In order to achieve this goal, the weight vectors must be modified so that they match the training set. The node k with the closest vector is that which gives the greatest input excitation **s** since this is just the dot product of the weight and input vectors. The weight vector of node k may be aligned more closely with the input if a change is made according to

$$\Delta w = \alpha(x - w) \qquad (8)$$

Competitive dynamics is useful in determining the node k, which is determined by the "winner-take-all" strategy. If y is the output of a node, where y is zero for all nodes except k, then the change in weights for each node can be shown as

$$\Delta w = \alpha(x - w)y \qquad (9)$$

Hence, the stages in learning (for a single vector presentation) are -

1) Apply vector at input to net and evaluate **s** for each node
2) Update the net (in practice, in discrete steps) according to (7) for a finite time or until it reaches equilibrium

3) Train all nodes according to (9)

If a network can learn a weight vector configuration like this, without being told about the nature of the input (existence of clusters, etc.) at the input, then it is said to undergo a process of *self-organized* or unsupervised learning. It is to be noted that the above simple example applies only to hard competition.

Adaptive Resonance Theory

Most learning techniques (including Competitive learning and Self-Organized maps) work only on static input environments. If a network is trained on a set of input vectors, it can classify the input environment correctly only if the input environment is not dynamic. In the presence of dynamically self-organizing data, the accuracy of a network (such as a back-propagation network or a Self-Organizing map) decreases rapidly because the fixed weights prevent the network from adapting to the changing environment. Thus, such networks are not *plastic*. To over come this problem. the networks can be retrained on a new set of input vectors. The network will adapt to any changes in the input environment but this causes a rapid decrease of accuracy with which it categorizes the old inputs because the old information is lost. Thus, this algorithm is not *stable*.

The above problem is called the stability-plasticity dilemma [Carpenter and Grossberg 1987a]. Adaptive Resonance Theory (ART) was designed specifically to solve this problem. The ART uses an incremental clustering algorithm to self-organize in real time and produce stable recognition while learning input patterns beyond those originally stored. ART incorporates feedback both at the level of the competitive network and between various modules at the network level. Both inhibitory and excitatory connections are used to achieve learning.

The simplest ART network is a vector classifier– it accepts as input a vector and classifies it into a category depending on the stored pattern it most closely resembles. Once a pattern is found, it is modified (trained) to resemble the input vector. If the input vector does not match any stored pattern within a certain tolerance, then a new category is created by storing a new pattern similar to the input vector. Consequently, no stored pattern is ever modified unless it matches the input vector within a certain tolerance.

ARTs have been applied to a variety of domains ranging from medical applications, such as classifying ECG patterns [Barro et. al. 1998], to associative memory and semantic data processing [Tan 1997]. ARTs have also found a niche in applications that require concept discovery, an example being automatic building of expert-systems [Ishihara et. al. 1995].

There are multiple versions of the ART – ART-1 [Carpenter and Grossberg 1987a] is a binary version that applies only to binary inputs, ART-2 [Carpenter and Grossberg 1987b] is an analog version that can cluster real valued inputs and the ART-3 [Carpenter and Grossberg 1990] uses ’chemical transmitters’ to control the search process in a hierarchical ART structure. In additions, many variations such as ARTMAP, Fuzzy ART, Gaussian ART, LAPART, PROBART and MART exist (for example, [Williamson 1996][Carpenter and Grossberg 1991]). In this paper we will confine ourself to a description of ART-1.

The ART-1 Network

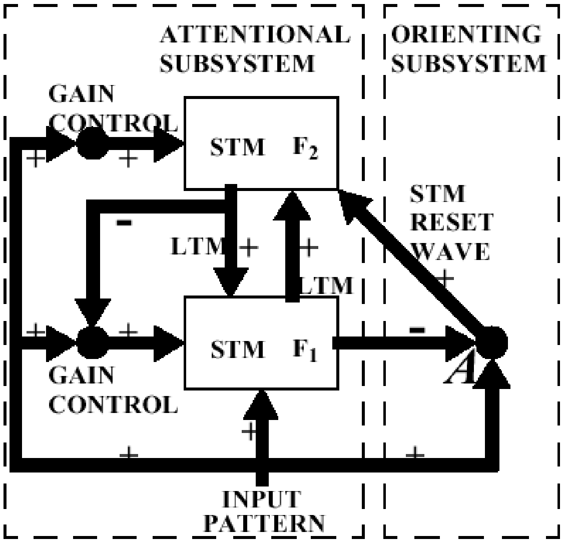


Fig. 4. ART-1 Network Model [Heins and Tauritz 1995]

ART-1 is the simplest ART learning model designed specifically to recognize binary patterns. The ART-1 system consists of an attentional sub-system and an orienting system as shown in Figure 4. The attentional subsystem consists of two competitive networks, the comparison layer F1 and the recognition layer F2, and two control gains.

The layers, F1 and F2, of the attentional subsystem encode patterns of activation in short-term memory (STM). Bottom-up and top-down pathways between F1 and F2 contain adaptive long-term memory (LTM) traces which multiply the signals in these pathways. The remainder of the circuit modulates these STM and LTM processes. F1 nodes are supraliminally activated (that is, sufficiently activated to generate output) if they receive a signal from at least two out of three possible input sources. The three are bottom-up input, top-down input and attentional gain control input. If a F1 node receives input from only one of these sources it is subliminally activated. This is called the 2/3 rule.

The orienting subsystem contains the reset layer for controlling the attentional subsystem's overall dynamics. The stabilization of learning and activation occurs in the attentional subsystem by matching bottom-up input activation and top-down expectation. The orienting controls the attentional subsystem when a mismatch occurs in the attentional subsystem. In other words, the orientating subsystem works as a novelty detector.

The ART works by hypothesis testing to check if the input vector can be classified into any currently existing cluster. If no such cluster exists, a new cluster is created for the current input, and stored in such a way that old information is not lost. The ART-1 hypothesis testing cycle consists of four steps [Heins and Tauritz 1995] -

1) Input pattern *I* generates the STM activity pattern *X* at F1 and activates both F1's gain control and the orienting subsystem A. Pattern *X* both inhibits A and generates the bottom-up signal pattern *S* which is transformed by the adaptive filter into the input pattern T. F2 is designed as a competitive network, only the node which receives the largest total input is activated (”winner-take-all”).
2) Pattern *Y* at F2 generates the top-down signal pattern *U* which is transformed by the top-down adaptive filter into the expectation pattern *V*. Pattern *Y* also inhibits F1's gain control, as a result of which only those F1 nodes that represent bits in the intersection of the input pattern *I* and the expectation pattern *V* remain supraliminally activated. If *V* mismatches *I* this results in a decrease in the total inhibition from F1 to A.
3) If the mismatch is severe enough A can no longer be prevented from releasing a nonspecific arousal wave to F2, which resets the active node at F2. The vigilance parameter ρ determines how much mismatch will be tolerated.
4) After the F2 node is inhibited its top-down expectation is eliminated and *X* can be reinstated at F1. The cycle then begins again. *X* once again generates input pattern *T* to F2, but a different node is activated. The previously chosen F2 node remains inhibited until F2's gain control is disengaged by removal of the input pattern.

The hypothesis testing cycle, which is a parallel search, repeats automatically at a very fast rate until one of three possibilities occurs -

- a F2 node is chosen whose top-down expectation approximately matches input *I*
- a previously uncommitted F2 node is selected
- the entire capacity of the system is used and input *I* cannot be accommodated

Until one of these outcomes prevails, no learning occurs as all STM computations proceed too quickly for any traces to be left in the LTM. Significant learning only occurs when the the hypothesis testing cycle ends and the system is in a *resonant* state.

The above algorithm only gives a qualitative, functional description of the ART-1 network. More implementation details and analysis using differential equations can be found in [Carpenter and Grossberg 1987a].

IV. Self-Organized Spatial Representations

A major portion of experiments in neuroscience has been directed towards mapping the brain i.e. finding the correspondence between portions of the brain and parts of the body they are concerned with. Another form of mapping has been to find the correspondence between groups of neurons and the feature of sensory input they react to. It has been established that this latter correspondence is spatial, meaning that groups of neurons dealing with similar features are also located near each other in the brain. In this section, we present self-organized models that could explain the development of such *brain maps*. Spatial organization of nodes in a neural network is also desirable in pattern classification problems. Self-Organized maps, which are essentially learning algorithms that produce spatially oriented neural networks, are also part of the focus of this section.

Maps in the Brain

Some of the most striking evidence for self-organization in the brain is the existence of various somatosensory, tonotopic, and retinotopic maps. In these maps, sheets or columns of neurons that respond to similar features of the sensory input are located near each other. In other words, the neurons organize themselves such that a topographic map of the skin on the body, sound frequencies, or visual features such as line orientation is created. The somatosensory map in the human cortex can be seen in Figure 5. Similarly, there are maps of body motion and the visual and auditory fields, some of which will be discussed later. This type of emergent pattern is not fixed or stable, and changes depending on the environment the animal is exposed to. For instance, the somatosensory maps of monkeys who had a finger amputated changed in such

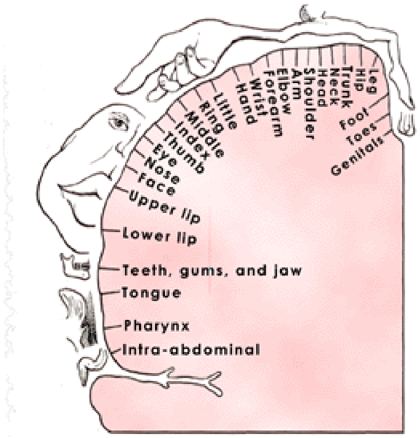


Fig. 5. The primary somatosensory area of the cortex is organized in a ”map” that is remarkably similar to the cartography of the body surface. This is called ”smooth” somatotopy. Notice, however, that the parts of the body used as organs of touch and feeling, such as the fingers and mouth, have disproportionately greater representation on the cortical map.

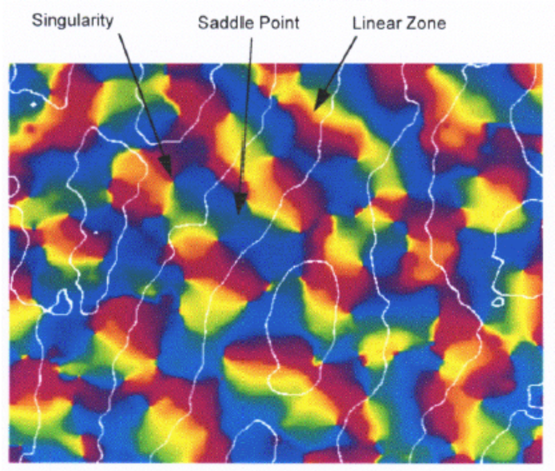


Fig. 6. Composite figure showing the arrangement of orientation domains (a single color represents a unique range of orientation preferences) and their relationship with ocular dominance column boundaries (white lines). The images were obtained by optical recording in macaque monkey striate cortex. Note that the iso-orientation domains tend to intersect ocular dominance column borders at right angles [Swindale 1996].

a way that the areas for the other fingers expanded into the region that was sensitive to the amputated finger. Not only do the local nearby regions expand, but the changes also propagate throughout the whole map so that many regions increase in size. This plasticity and reorganization occurs even in adult brains. This suggests that a template or recipe is not being used, and the process might be self-organized. Added to this, there have been many models using Hebbian learning as a basis which can account for a great deal of the features present in these maps.

Self-Organization in the Visual Cortex

The most well-studied type of sensory maps in the brain are those located in the visual cortex. The primary visual cortex, as discussed in the previous section, is a topographic map in which adjacent neurons respond to adjacent regions of the visual field. In addition, different neurons respond more strongly to particular features of the input such as ocularity (which eye the input came from), line orientation, size, and temporality. In general, neurons with the same feature preferences group themselves together into columns. If electrodes are moved vertically through the thickness of the cortex, it has been found that most neurons have the same orientation and ocularity preferences. If the electrodes are moved tangentially through the cortex, the cells first respond to left eye inputs, then both, then right eye, then both, then left eye, etc. If the electrodes are moved tangentially in the orthogonal direction, it was found that the neurons are selective for vertical inputs, then diagonal, then horizontal, etc. In Figure 6, the patterns of selectivity that result can be seen. Each color represents a range of orientation preferences, while the white lines represent the borders of the ocular dominance columns. As can be seen from the figure, the maps are not perfectly continuous. There are points around which the orientation preference changes continuously in a circular fashion, called singularities or pinwheels. Between adjacent singularities there are regions where the orientation preference changes slowly and continuously in an approximately linear fashion, called linear zones. There are other regions between singularities where local minima of orientation preference in orthogonal directions exist, called saddle points. It is also worth noting that ocularity and orientation patterns are not independent. As can be seen from Figure 6, the lines representing regions of different ocularity are somewhat orthogonal to the elongated ellipse shapes representing regions of similar orientation.

The neuronal specialization that occurs is not hardwired or fixed. For example, kittens were found to have abnormal orientation selection among neurons when they were blinded at birth. Also, it has been found that even an adult cortex can undergo reorganization of the connections when experiencing sensory or cortical manipulation such as lesions. However, evidence suggests that some specialization does occur before structured visual stimuli is experienced, and in some animals exists in some form before birth. Later introduction

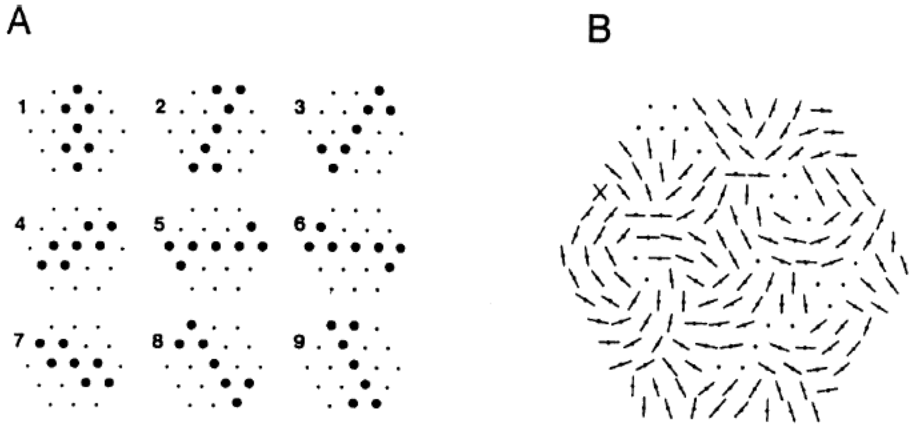


Fig. 7. (A) The nine different stimuli used by von der Malsburg. The large dots are active units. (B) The layout of orientation preference in the simulated cortex after 100 learning steps. The dots represent units which failed to learn to respond to any of the stimuli [Swindale 1996].

of structured visual input skews the distribution of orientation selectivity. A possible solution to this dilemma was proposed by Ralph Linsker and others that followed him, who demonstrated that a simple feed-forward network that used Hebbian learning could develop orientation preference similar to that seen in the visual cortex even given unstructured visual input. Unstructured visual input can be spontaneously occurring retinal activity, which some have even suggested might be internally-generated input patterns encoded in genetic information [Bednar and Miikkulainen 2003]. In order to account for the various patterns that emerge in the visual cortex many models, most of which use self-organization, have been proposed. A comparison of some models including an analysis of the similarity that exists in their underlying assumptions can be found in [Erwin et. al. 1995] [Swindale 1996]. Most models are based on similar assumptions [Swindale 1996]:

1) Hebb synapses;
2) Correlated or spatially patterned activity in the afferents to cortical neurons;
3) Fixed connections between cortical neurons which are locally excitatory and inhibitory at slightly greater distances
4) Normalization of synapse strength.

In the following subsections, we will discuss the details of two models in order to give an idea of what they look like.

Von der Malsburg's model

One of the first correlation-based models was described by Von Der Malsburg [von der Malsburg 1973]. His model first assumed a Hebbian rule that is competitive (but in a simpler manner than described before). Here, competitive refers to the fact that an increase in strength to one neuron results in a relative decrease to another. This was enforced using normalization, which keeps the total strength of incoming connections to a neuron constant. This competition results in neurons that respond to correlated inputs. If two inputs are activated by correlated activity, then they mutually reinforce their connections since they both work together to activate the target cell. This type of positive feedback was discussed earlier in this paper.

Von der Malsburg had two types of cells: cortical cells (both inhibitory and excitatory) and retinal cells, representing a two layered network (although one layer is just the input given to the system). There were a fixed number of excitatory and inhibitory cells connected together in a fixed manner such that the inhibitory connections went slightly farther than excitatory, which is something that is thought to occur in real animal brains. Each excitatory cell had connections to all 19 retinal input neurons and each retinal neuron was connected with every excitatory cell, whose weights were random at first. The following non-linear differential equation was used in order to model the changes in response of the cells:

$$\frac{dH_k(t)}{dt} = -\alpha_k H_k(t) + \sum_{l=1}^{N} p_{lk} H_l^*(t) + \sum_{i=l}^{M} s_{ik} A_i^*(t) \qquad (10)$$

$$k = 1, \ldots, N$$

The first term on the right hand side of the equation represents the decay of a neuron with time. α_k is the decay constant and $H_k(t)$ is the response of the cell at time t. The second term represents the excitation and inhibition from other lateral cells. p_{lk} is the connection

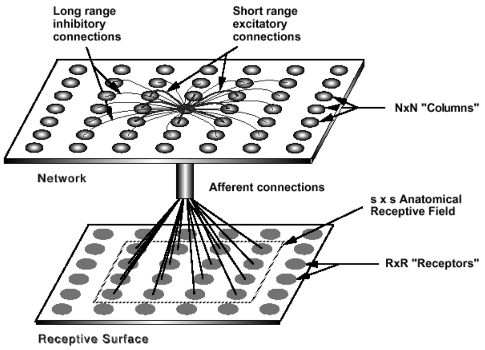


Fig. 8. The RF-LISSOM model. The lateral excitatory and lateral inhibitory connections of a single neuron in the network are shown, together with its afferent connections [Miikkulainen et. al. 1997].

strength between cell l and cell k and $H_l^*(t)$ is the value of $H_l(t)$ after the application of a threshold function. The last term is the effect of the retinal cells, where s_{ik} is the strength of the connection and $A_i^*(t)$ is the stimulus strength of retinal cell i after applying a threshold function to it. The rule for changing the weights of the input connections (between retinal and cortical cells) was as follows:

$$\Delta s_{ik} \propto A_i^* H_k^* \tag{11}$$

After each learning step, the total of the synaptic strengths coming into each cortical cell was normalized by multiplying it by a factor proportional to $1/\sum_i s_{ik}$. Note that p_{lk}, the weights between lateral cells were not changed and remained fixed. Figure 7 shows the inputs given to the system and the resulting map which demonstrates orientation preference. This model was the first to produce this pattern, and the first to use local connectivity that had short range excitatory connections and long range inhibitory connections in a sheet of cells.

The model was later extended by Malsburg and others in order to account for retinotopy, orientation columns, and ocular dominance.

RF-LISSOM

It is important to distinguish lateral and afferent connections existing in neuronal structures. As shown in Figure 8, lateral connections are the inhibitory and excitatory connections among neurons, and afferent connections are connections among different layers (for example between retinal and cortical cells in Malsburg's model). In most models, the lateral connections are assumed to be fixed with the excitatory connections being shorter range than inhibitory connections. One model, called RF-LISSOM (Receptive Field Laterally Interconnected Synergetically Self-Organizing Maps) develops both afferent and lateral connections together, and can account for orientation, ocular dominance, and size selectivity columns as well as low-level phenomena such as tilt aftereffects [Miikkulainen et. al. 1997].

The RF-LISSOM model uses layers of sheets of interconnected neurons (as seen in Figure 8). Through afferent connections each neuron receives input from a receptive surface and also has reciprocal lateral inhibitory and excitatory connections. As before, lateral inhibitory connections run for longer distances than excitatory. This is related to the Self-Organizing Maps used in artificial intelligence, discussed later. Each neuron in the N x N network receives input from a receptive field of size s x s where s is half the size of the retina (the receptive surface). The weights of both afferent and lateral connections have positive weights, start out at random, and change using similar Hebbian rules. The initial response n_{ij} of neuron (i,j) is calculated as:

$$\eta_{ij} = \sigma\left(\sum_{r_1,r_2} \xi_{r_1,r_2}\mu_{ij,r_1r_2}\right) \tag{12}$$

Here ξ_{r_1,r_2} is the activation level of the retinal receptor (r_1,r_2) within the receptive field of the neuron, μ_{ij,r_1r_2} is the strength of the afferent connection, and σ is a piecewise linear approximation of the sigmoid activation function. At each step, the response of the neurons changes as follows:

$$\eta_{ij}(t) = \sigma\Big(\sum_{r_1,r_2} \xi_{r_1,r_2}\mu_{ij,r_1r_2} + \gamma_e \sum_{k,l} E_{ij,kl}\eta_{kl}(t-\delta t) - \gamma_i \sum_{k,l} I_{ij,kl}\eta_{kl}(t-\delta t)\Big) \tag{13}$$

The first term is the afferent connection, the second term is the effect of the excitatory connections, and the third term is the effect of the inhibitory connections. γ_e and γ_i are just scaling factors determining how much lateral excitatory and inhibitory interaction is desired. At first, the activity on the receptive field is spread out across the map, but after a few iterations it will stabilize in a patch of activity. After this, the weights of the connections (both afferent and lateral) are modified by the following rule:

$$w_{ij,mn}(t+1) = \frac{w_{ij,mn}(t) + \alpha\eta_{ij}X_{mn}}{\sum_{mn}\left[w_{ij,mn}(t) + \alpha\eta_{ij}X_{mn}\right]} \tag{14}$$

This is basically a normalized Hebb rule. Both the inhibitory and excitatory lateral connections are strengthened by correlated activity, and since correlated activity is rare among neurons far away from each other the long range connections eventually become weak. These weak connections are manually pruned by the system periodically if they become too weak, reflecting evidence that such connections die early in development in animals. The radius of lateral excitatory interactions starts out large but is decreased until it covers only nearest neighbors. As the system goes through many iterations, the activity pattern decreases in radius and eventually converges to small activity areas called “activity bubbles”. A more detailed explanation and many simulation results can be found in [Miikkulainen et. al. 1997]. Although more complicated than other models, it is able to account for a wide range of patterns and correlations that occur in real animals. This model and others like it use a class of more general mechanisms called Self-Organizing Feature Maps, which are discussed in the next section.

Self-Organizing Feature Maps

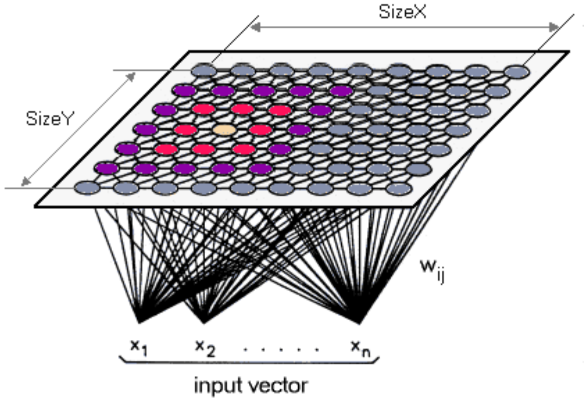


Fig. 9. Activation in a Kohonen Network [Kohonen 1990]

Self-Organizing Feature Maps (SOFMs) [Kohonen 1982], also called Kohonen maps or Kohonen networks after their inventor, are neural networks that effectively project input space on prototypes of low dimensional regular grids. This property makes them extremely useful in visualization and exploration of data properties. Another important property of Kohonen networks is that they preserve the topology of the input space during mapping. A Kohonen network implements soft competitive learning in a sheet-like artificial neural network (Figure 9), the cells of which become specifically tuned to various input signal patterns or classes of patterns through an unsupervised learning process. In the basic version, only one cell or a local group of cells at a time gives the active response to the current input. The locations of the responses tend to become ordered as if some meaningful coordinate system for different input features were being created over the network. The spatial location or coordinates of a cell in the network then correspond to a particular domain of input signal patterns.

Kohonen maps bear strong physiological links to the structure of the human brain and the working of a Kohonen network has been shown to be biologically plausible [Kohonen 1993]. Many brain maps, especially those in primary sensory areas, are ordered according to some feature dimensions of sensory signals. These include the tonotopic maps of the auditory cortex and the somatotopic map discussed earlier. Some of these maps represent abstract qualities of sensory and other experiences, an example being the organization of neural responses according to categories and semantic values of words (cf. [Caramazza 1988]). It thus seems that internal representations of information in the brain are generally organized spatially, although there is only partial biological evidence for this.

The Self-Organizing Map Algorithm

The Kohonen network uses a soft competitive learning algorithm as explained here (from [Kohonen 1990]). Let $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T \in \Re_n$ be the input vector (in matrix notation). Thus input is connected in parallel to all the neurons i in the network. The weight vector of cell i is $\mathbf{m_i} = [\mathbf{m_{i1}}, \mathbf{m_{i2}}, \ldots, \mathbf{m_{in}}]^T \in \Re_n$. The simplest analytical measure for the match of $\mathbf{x}$ with $\mathbf{m_i}$ is the inner product $\mathbf{x^T m_i}$. However, the euclidean distance between $\mathbf{x}$ and $\mathbf{m_i}$ can also be used, in which case the winner is the node with the least distance. Lateral interaction is enforced in a general form by defining a neighborhood set N_c around a cell c. At each learning step, all the cells within N_c are updated while those outside N_c are left intact. The neighborhood is centered around the cell for which the best match with input $\mathbf{x}$ is found (the winner in a “winner-take-all” network) :

$$\parallel \mathbf{x} - \mathbf{m_c} \parallel = min_i \{ \parallel \mathbf{x} - \mathbf{m_c} \parallel \} \qquad (15)$$

The width or radius of N_c can be time-variable. For a good global ordering, N_c should initially be large (up to half the size of the network) and should shrink monotonically with time. This may be explained by observing that an initial coarse spatial resolution in the learning process, that induces rough global order, can be followed by finer adjustments. The updation process can be expressed as -

$$\mathbf{m_i}(t+1) = \begin{cases} \mathbf{m_i}(t) + \alpha(t)[\mathbf{x}(t) - \mathbf{m_i}(t)] & \text{if i} \in N_c(t) \\ \mathbf{m_i}(t) & \text{if i} \notin N_c(t) \end{cases} \tag{16}$$

where $\alpha(t)$ is a scalar valued "adaptation gain" $0 < \alpha(t) < 1$. This is the Hebbian learning rule for the competitive network. Alternatively, this may be accomplished by introducing a scalar "kernel" function $h_{ci} = h_{ci}(t)$,

$$\mathbf{m_i}(t) = \mathbf{m_i}(t) + h_{ci}(t)[\mathbf{x}(t) - \mathbf{m_i}(t)] \tag{17}$$

The definition of h_{ci} can be general, one of the commonly used ones being the "mexican hat" waveform shown in Figure 10. This function encodes long-range inhibitory and short-range excitatory connections that are similar to neuron interactions in the brain.

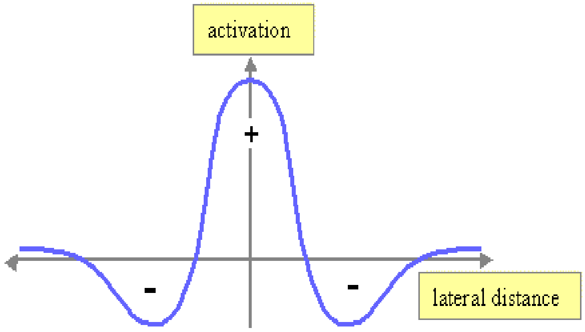


Fig. 10. Mexican Hat Neighborhood Function

The Kohonen learning algorithm is then simply the updation of the weights using equation 17 on the neighborhood centered on the winner of the competition. This simple algorithm provides spatial ordering of any input domain where a set of features of interest are defined.

Ordering in Kohonen networks

This section illustrates the working of Kohonen networks through some demonstrations of the ordering process. The examples highlight the effect of the weight vectors approximating the the density function of the input vectors in an orderly manner. The input vectors $\mathbf{x_i}(t)$ are drawn from this density function independently and at random, after which they cause adaptive changes in the weight vectors $\mathbf{m_i}$. The $\mathbf{m_i}$ appear as points in the same coordinate system as that in which the $\mathbf{x_i}(t)$ are represented; in order to indicate to which unit each $\mathbf{m_i}$ value belongs, the points corresponding to the $\mathbf{m_i}$ array have been connected by a grid conforming to the topology of the processing unit array. In other words, the weight vectors $\mathbf{m_i}$ and $\mathbf{m_j}$ are connected only if the corresponding units i and j are adjacent in the array. In Figure 11(a), the arrangement of cells is rectangular, while in Figure 11(b), the cells are connected in a linear chain. The figures also show the intermediate phases during the self-organization. Figure 11(c) shows the application of the Kohonen network to the Travelling Salesman Problem. The network configuration is a closed chain of 100 nodes. The tendency of the network to preserve neighborhood results in attempting short tours as the final configuration. In the figure, the network finds the shortest possible tour, though this is not always the case.

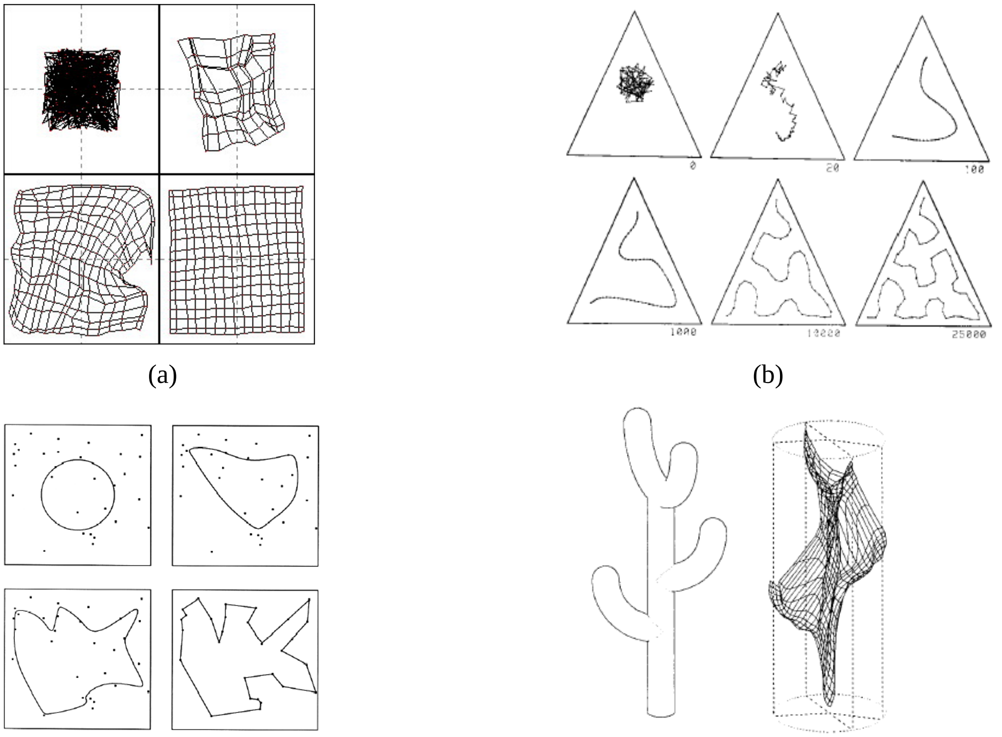
The results are more interesting if the dimensionalities of the input and weight vectors differ, as in Figures 11(b) and 11(d). In the latter figure, a two-dimensional grid approximates a three-dimensional probability grid.

Applications and Variations

Self-organized networks have been applied to pattern recognition [Kohonen and Somervuo 1998], image analysis [Villman and Mernyi 2001], process-control [Kasslin et. al. 1992], data mining and analysis [Honkela et. al. 1996], analysis of financial statements [Deboeck 1998], to name a few. Similarly, a large number of variations and improvements on the basic learning algorithm exist. These include various methods of finding the winning node and different methods of defining the neighborhood function.

While the Kohonen network uses unsupervised learning, "fine-tuning" of the learning process and its speed-up may be accomplished by using supervised learning. Algorithms for this purpose are called *Learning Vector Quantization (LVQ)* algorithms [Kohonen 1990]. These are not discussed here as they do not involve self-organization.

Since the learning process is stochastic, the final statistical accuracy of the mapping depends on the number of steps, which must be reasonably large. Typically, about 100000 steps are required, though some applications may require as few as 10000 steps. Also, if the neighborhood is too small to begin with, the map degenerates into various mosaic-like partitions with no global ordering. A phase-transition map based on the behavior of the network with respect to the variation in neighborhood size can be built analytically for the simplest cases, but requires lengthy mathematical analysis, and hence is not presented here. Behavior varies from a uniform distribution across nodes to sharply clustered outputs with sharp phase transitions among the various phases [Graepel et. al. 1997].



(a)

(b)

(c)

(d)

Fig. 11. Evolution of Kohonen networks in different applications [Kohonen 1990]: (a) Weight vectors during ordering: 2D array. (b) Weight vectors during ordering: 1D array. (c) Solving the TSP problem for 30 cities: The iterations shown are 5000, 7000, and 10000. (d) Representation of a 3D array by a 2D array.

V. Discussion

There has been a great deal of neuroscientific and theoretical research investigating the mechanisms that enable the brain. Self-organization is a likely candidate and has been successfully applied to model the brain. In fact, most of the successful models use Hebbian learning, in which neurons use only local information. These models have many of the characteristics needed in self-organized systems. One example is the feedback that occurs when correlated inputs are fed into a Hebbian learning system, as discussed in this paper previously. Other possible mechanisms seem unlikely, usually simply because of the complexity and plasticity of the brain. We will now discuss the other possible mechanisms and argue why they probably cannot account for the full functionality of the brain. According to [Camazine et. al. 2001], the four other forms of organization are: leader, blueprint, recipe, template.

One ”leader” neuron or section of the brain probably cannot keep track of the billions of neurons in the brain, and such a scheme would leave even a small amount of brain damage fatal if it was in the right place. This does not, however, rule out a hierarchical system of leadership when looked at from a larger scale. For instance, the brain has been shown to be made up of somewhat specialized regions that do particular things well. This has been discovered through imaging of the brain (eg. using fMRI) to see which parts are active during specific tasks. One well known example is that the hippocampus has been found to be involved in the functioning of memory. Looking at it from this perspective does not rule out self-organization, since the parts themselves are made of a great deal of neurons that are left to be explained for. It is entirely possible that the self-organizing process results in the specialization of various parts of the brain; that could be the pattern that emerges. There is some precedence for this in the visual cortex, since individual neurons develop specializations to different features of the input.

DNA can also possibly act as a blueprint for the brain, but it does not have enough information to account for all of the connections in the brain from an information theoretic point of view. Furthermore, it is well known that everyone has a different brain especially when it is developed or experiencing different environments. This is true even for identical twins who have the same DNA. The information contained in DNA is unlikely to contain contingencies for all possible environments the brain might encounter. Another method, a recipe, is also unlikely for the same reasons: It would encounter the same information size problem.

It is not obvious at first whether the brain could be formed from a template. One definition of a template is as follows [Camazine et. al. 2001]:

> "A full size guide or mold that specifies the final pattern and strongly steers the pattern formation"

It can be said that the environment (which includes the body of the animal itself) "strongly steers the pattern formation". The various somatosensory maps such as those found in the visual cortex reorganize themselves depending on the visual input it is given throughout its lifetime (especially during development but even during adulthood). The final pattern, however, is certainly not encoded or specified in the organism itself. It is also not necessarily exactly the same shape or size, although the relative sizes might have some strong correlation. This is certainly true for other examples of self-organization since the environment ants live in shapes the pattern, but is not located inside the organism itself. Hence, this form of influence does not seem to fit this specific definition of a template.

The success of models of visual cortex development provides some clue that self-organization might be the process causing these patters in the brain, although it is not certainly not hard proof. It is also likely that self-organization arises in many forms; different Hebbian rules could be used in the brain, with different patterns and results emerging. Also, there is some evidence that genetically determined eye-brain and fiber-fiber markers are needed to control the activity dependent self-organizing process in order to account for various experimental data found in animals [Cowan and Friedman 1995]. Hence even though none of the other organizing principles can account for the complete brain, this suggests that not only one mechanism is involved. Since the brain is such a complex system, it could be that all of the various organizing methods are used, including self-organization. Fortunately, advances in technology used by neuroscientists might allow a greater understanding of these processes in the future.

Self-organized artificial intelligent systems were designed to specifically incorporate mechanisms that are active in the brain. However, in keeping with the different application needs, there has been a divergence between biology and AI. While feedback mechanisms exist at different levels in the neural network structures discussed previously, not all of these lead to self-organized behavior. In Competitive learning, for example, feedback occurs at both the neuron (or node) interaction level, and during modification of weights. Feedback through lateral inter-neuronal connections leads the emergence of a 'winning' node solely through local interactions and constitutes self-organized behavior. However, the modification of weights is done with the aim of driving the weight vectors closer to the input. Here each node receives the input vector and hence, this constitutes global knowledge. The changes in the system are also made keeping in mind the final goal to be achieved, i.e. a matching of input and weight vectors. Similarly, in the ART-1 network discussed previously, feedback occurs in the F1 and F2 modules that are competitive networks. Mutual positive feedback is also active between the attentional and orientational sub-systems and leads to 'resonance'. which in turn results in learning. The learning, however, is not self-organized in that the resonance state is not brought about by any self-organized mechanism operating at the module level of the network.

Kohonen networks achieve topological ordering through a competitive network that has a neighborhood function associated with weight changes. As before, self-organization occurs at the neuronal level. In addition, the network also organizes itself topologically relative to the input through only local interactions. The weight adaptations in local (neighborhood) regions leads to a spatially grouped representation of the input. A consequence of using such a mechanism is that the accuracy of the network increases monotonically both with increasing number of nodes and with greater learning experience [Goppert and Rosenstiel 1993]. This is not the case with, say, back-propagation networks that can suffer from *over-learning*, where the network accuracy decreases with more learning.

In the AI domain, unsupervised learning is also referred to as self-organized learning. However, this does not necessarily imply any self-organized mechanism at work. It is self-organized in the sense that no output is produced that specifies the direction in which the network adaptation should be performed. In contrast, back-propagation networks produce an error signal on which gradient descent is executed to achieve learning. Thus, while unsupervised learning methods exist that are truly self-organized (such as Competitive learning), there also exist Bayesian probabilistic learning methods that are unsupervised but not self-organized.

VI. REFERENCES

[Barro et. al. 1998] Barro, S., M. Fernandez-Delgado, J.A. Vila-Sobrino, C.V. Regueiro, and E. Sanchez. 1998. "Classifying multichannel ecg patterns with an adaptive neural network". *IEEE*

Engineering in Medicine and Biology Magazine, **17**(1):45–55.

[Bednar and Miikkulainen 2003] Bednar, J.A. and R. Miikkulainen. 2003 (in press). “Learning Innate Face Preferences”. *Neural Computation*, **15**(7).

[Brown and Chattarji 1995] Brown, T.H. and S. Chattarji. 1995. “Hebbian Synaptic Plasticity”. Pages 454–459 in: *The Handbook of Brain Theory and Neural Networks*, (M. Arbib, Ed.), MIT Press, Cambridge, Mass.

[Calvin 1995] Calvin, W.H. 1995. “Cortical Columns, Modules, and Hebbian Cell Assemblies”. Pages 269–272 in: *The Handbook of Brain Theory and Neural Networks*, (M. Arbib, Ed.), MIT Press, Cambridge, Mass.

[Camazine et. al. 2001] Camazine, S., J. L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau. 2001. “Self-organization in Biological Systems”. Princeton University Press, Princeton, NJ.

[Caramazza 1988] Caramazza, A. 1988. “Some aspects of language processing revealed through the analysis of acquired aphasia: The lexical system”. *Annual Review of Neuroscience*, **11**:395–421.

[Carpenter and Grossberg 1987a] Carpenter, G., and S. Grossberg. 1987. “A massively parallel architecture for a self-organizing neural pattern recognition machine”. *Computer Vision, Graphics and Image processing*, **37**:54–115

[Carpenter and Grossberg 1987b] Carpenter, G., and S. Grossberg. 1987. “ART 2: Self-organization of stable category recognition codes for analog input patterns”. *Applied Optics*, **26**(23):4919–4930

[Carpenter and Grossberg 1990] Carpenter, G., and S. Grossberg. 1990. “ART-3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures”. *Neural Networks*, **3**:129–152

[Carpenter and Grossberg 1991] Carpenter, G.A., S. Grossberg, and D.B. Rosen. 1991. “Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system”. *Neural Networks*, **4**:759–771.

[Cowan and Friedman 1995] Cowan, J.D. and A.E. Friedman. 1995. “Development and Regeneration of Eye-Brain Maps”. Pages 295–299 in: *The Handbook of Brain Theory and Neural Networks*, (M. Arbib, Ed.), MIT Press, Cambridge, Mass.

[Deboeck 1998] Deboeck, G. 1998. “Financial Applications of Self-Organizing Maps”. *Neural Network World*, **8**(2):213–241.

[Erwin et. al. 1995] Erwin, E., K. Obermayer, and K. Schulten. 1995. “Models of orientation and ocular dominance columns in the visual cortex: A critical comparison”. *Neural Computation*, **7**:425–468.

[Fukushima 1975] Fukushima, K. 1975. “Cognitron: A self-organizing multi-layered neural network”. *Biological Cybernetics*, **20**:121–136.

[Fukushima 1980] Fukushima, K. 1980. “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”. *Biological Cybernetics*, **36**:193–202.

[Fyfe 1997] Fyfe C. 1997. “A Neural Network for PCA and Beyond”. *Neural Processing Letters*, **6**(1):1–9.

[Gazzaniga et. al. 2002] Gazzaniga, M.S., R. Ivry, and G.R. Mangun. 2002. “Cognitive Neuroscience”. W.W. Norton, 2nd Edition.

[Gerstner and Kistler 2002] Gerstner, W., and W. Kistler. 2002. *Spiking Neuron Models*, Cambridge University Press.

[Goppert and Rosenstiel 1993] Goppert, J. and W. Rosenstiel. 1993. “Self-organizing maps vs. backpropagation: An experimental study”. *Proc. of Workshop on Disign Methodologies for Microelectronis and Signal Processing*, pp. 153–162.

[Graepel et. al. 1997] Graepel, T., M. Burger, and K. Obermayer. 1997. “Phase transitions in stochastic self-organizing maps”. *Physical Review E*, **56**(4):3876–3890.

[Grossberg 1976a] Grossberg, S. 1976. “Adaptive pattern classification and universal recording, 1: Parallel development and coding of neural feature detectors”. *Biological Cybernetics*, **23**:121–134.

[Grossberg 1976b] Grossberg, S. 1976. “Adaptive pattern classification and universal recording, 2: Feedback, expectation, elfaction and ellusions”. *Biological Cybernetics*, **23**:187–202.

[Heins and Tauritz 1995] Heins, L.G. and D.R. Tauritz. 1995. “Adaptive Resonance Theory (ART): An Introduction”. *Technical Report 95-35*, Leiden University.

[Honkela et. al. 1996] Honkela, T., S. Kaski, K. Lagus, and T. Kohonen. 1996. “Exploration of full-text databases with self-organizing maps”. *Proceedings of ICNN’96, IEEE International Conference on Neural Networks*, **I**56–61.

[Ishihara et. al. 1995] Ishihara, K., S. Ishihara, Y. Matsubara, M. Nagamachi. 1995. “An Automatic Builder for a Kansei Engineering Expert System Using Self- Organizing Neural Networks”. *International Journal of Industrial Ergonomics*, **15**(1)13–24.

[Kaplan 1991] Kaplan,S., M. Weaver, and R.M.

French. 1991. “Active Symbols and Internal Models: Towards a Cognitive Connectionism”. *AI and Society* **4**(1):51–71.

[Kasslin et. al. 1992] Kasslin, M., J. Kangas, and O. Simula. 1992. “Process state monitoring using self-organizing maps”. *Articial Neural Networks*(I. Aleksander and J. Taylor, Eds.), **2**(2):1531–1534, North-Holland.

[Kelso 1995] Kelso, J. A. S. 1995. “Dynamic Patterns: The Self-Organization of Brain and Behavior”. MIT Press, Cambridge, Mass.

[Kohonen 1982] Kohonen, T. 1982. “Self-organized formation of topologically correct feature map”. *Biological Cybernetics*, **43**:59–69.

[Kohonen 1990] Kohonen, T. 1990. “The self-organizing map”. *Proceedings of IEEE*, **78**:1464–1480.

[Kohonen 1993] Kohonen, T. 1993. “Physiological interpretation of the self-organizing map algorithm”. *Neural Networks*, **6**:895–905.

[Kohonen and Somervuo 1998] Kohonen, T. and P. Somervuo. 1998. “Self-Organizing Maps of Symbol Strings”. *Neurocomputing*, **21**:19–30.

[Martinetz and Schulten 1991] Martinetz, T.M. and K. J. Schulten. 1991. “A “neural-gas” network learns topologies”. Pages 397–402 in: *Artificial Neural Networks* (T. Kohonen, K. Mkisara, O. Simula, and J. Kangas, Eds.), North-Holland, Amsterdam.

[McCulloch and Pitts 1943] McCulloch,W.S. and W. Pitts. 1943. “A Logical Calculus of the Ideas Immanent in Nervous Activity”. *Bulletin of Mathematical Biophysics* **5**:113–133.

[McDermott 1996] McDermott, J. 1996. “The Emergence of Orientation Selectivity in Self-Organizing Neural Networks”. *The Harvard Brain*, **3**(1):43–51

[Merzenich et. al. 1988] Merzenich, R.J., G. Recanzone, W.M. Jenkins, T.T. Allard and R.J. Nudo. 1988. “Cortical representation plasticity”. Pages 41–68 in: *Neurobiology of Neocortex* (P. Rakic and W. Singer, Eds.), Wiley, New York.

[Miikkulainen et. al. 1997] Miikkulainen, R., J. A. Bednar, Y. Choe, and J. Sirosh. 1997. “Self-organization, plasticity, and low-level visual phenomena in a laterally connected map model of the primary visual cortex”. Pages 257–308 in: *Perceptual Learning*, vol. 36 of *Psychology of Learning and Motivation*, (R.L. Goldstone, P.G. Schyns, and D.L. Medin, Eds.), Academic Press, San Diego, CA.

[Miller 1996] Miller, K.D. 1996. “Receptive Fields and Maps in the Visual Cortex: Models of Ocular Dominance and Orientation Columns”. Pages 55–78 in: *Models of Neural Networks III*, (E. Domany, J.L. van Hemmen, and K. Schulten, Eds.), Springer-Verlag, NY.

[Nass and Cooper 1975] Nass, N.M. and L.N. Cooper. 1975. “A theory for the development of feature-detecting cells in visual cortex”. *Biological Cybernetics*, **19**:1–18.

[Nowlan 1991] Nowlan, S.J. 1991. “Soft competitive adaptation: Neural network learning algorithms based on fitting statistical mixtures”. PhD Dissertation. Carnegie-Mellon University.

[Oja 1982] Oja E. 1982. “A simplified neuron model as a principal component analyser”. *Journal of Mathematical Biology*, **16**:267–273.

[Oja 1989] Oja E. 1989. “Neural networks, principal components and subspaces”. *International Journal of Neural Systems*, **1**:61–68.

[Perez et. al. 1975] Perez, R., L. Glass, and R.J. Shlaer. 1975. “Development of specificity in cat visual cortex”. *Journal of Mathematical Biology*, **1**:275–288.

[Ramoa et. al. 1986] Ramoa, A.S., M. Shadlen, B.C. Skottun, R.D. Freeman. 1986. “A comparison of inhibition in orientation and spatial frequency selectivity of cat visual cortex”. *Nature*, **321**:237–239.

[Ruf and Schmitt 1998] Ruf B. and M. Schmitt. 1998. “Self-organization of spiking neurons using action potential timing”. *IEEE Transactions on Neural Networks*, **9**(3):575–578.

[Rumelhart et. al. 1986] Rumelhart, D.E., G.E. Hilton, and R.J. Williams. 1986. “Learning internal representations by error propagation”. Pages 318–362 in: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1: Foundations* (D.E. Rumelhart and J. McClelland, Eds.), M.I.T. Press, Cambridge, MA.

[Tan 1997] Tan, A. 1997. “Cascade artmap: Integrating neural computation and symbolic knowledge processing”. *IEEE Transactions on Neural Networks*, **8**(2):237–250.

[Swindale 1996] Swindale, N.V. 1996. “The development of topography in the visual cortex: a review of models”. *Network* **7**:161–247.

[Villman and Mernyi 2001] Villmann, T. and E. Mernyi. 2001. “Extensions and Modifications of the Kohonen-SOM and Applications in Remote Sensing Image Analysis”. Pages 121–145 in: *Self-Organizing Maps: Recent Advances and Applications* (U.Seiffert and L.C. Jain, Eds.), Springer-Verlag.

[von der Malsburg 1973] von der Malsburg, C. 1973. "Self-organization of orientation sensitive cells in the striate cortex". *Kybernetik*, **14**:85–100.

[von der Malsburg 1995] von der Malsburg, C. 1995. "Self-organization and the brain". Pages 840–843 in: *The Handbook of Brain Theory and Neural Networks*, (M. Arbib, Ed.), MIT Press, Cambridge, Mass.

[Widrow and Hoff 1960] Widrow, B. and Hoff, M. 1960. "Adaptive switching circuits". *IRE WESCON Convention Record*, **4**:96–104.

[Williamson 1996] Williamson, J.R. 1996. "Gaussian artmap: A neural network for fast incremental learning of noisy multidimensional maps". *Neural Networks*, **9**(5):881–897.