ABSTRACT

A neural network which self-organizes and self-stabilizes its recognition codes in response to arbitrary orderings of arbitrarily many and arbitrarily complex binary input patterns is here outlined. Top-down attentional and matching mechanisms are critical in self-stabilizing the code learning process. The architecture embodies a parallel search scheme which updates itself adaptively as the learning process unfolds. After learning self-stabilizes, the search process is automatically disengaged. Thereafter input patterns directly access their recognition codes, or categories, without any search. Thus recognition time does not grow as a function of code complexity. A novel input pattern can directly access a category if it shares invariant properties with the set of familiar exemplars of that category. These invariant properties emerge in the form of learned critical feature patterns, or prototypes. The architecture possesses a context-sensitive self-scaling property which enables its emergent critical feature patterns to form. They detect and remember statistically predictive configurations of featural elements which are derived from the set of all input patterns that are ever experienced. Four types of attentional process—priming, gain control, vigilance, and intermodal competition—are mechanistically characterized. Top-down priming and gain control are needed for code matching and self-stabilization. Attentional vigilance determines how fine the learned categories will be. If vigilance increases due to an environmental disconfirmation, then the system automatically searches for and learns finer recognition categories. A new nonlinear matching law (the 2/3 Rule) and new nonlinear associative laws (the Weber Law Rule, the Associative Decay Rule, and the Template Learning Rule) are needed to achieve these properties. All the rules describe emergent properties of parallel network interactions. The architecture circumvents the saturation, capacity, orthogonality, and linear predictability constraints that limit the codes which can be stably learned by alternative recognition models.

† Supported in part by the Air Force Office of Scientific Research (AFOSR 85-0149 and AFOSR F49620-86-C-0037) and the National Science Foundation (NSF DMS-84-13119).

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR 85-0149 and AFOSR F49620-86-C-0037), the Army Research Office (ARO DAAG-29-85-K0095), and the National Science Foundation (NSF IST-84-17756).

Acknowledgement: We wish to thank Carol Yanakakis for her valuable assistance in the preparation of the manuscript.
SEARCH CYCLE:
INTERACTIONS BETWEEN ATTENTIONAL
AND ORIENTING SUBSYSTEMS

The neural network outlined herein is called an ART system, after the adaptive resonance theory introduced by Grossberg\textsuperscript{1}. More recently, ART networks have been further characterized, and their dynamic properties have been derived in a series of theorems\textsuperscript{2-4}. A single cycle of the search process carried out by this ART network is depicted in Figure 1. In Figure 1a, an input pattern I generates a short term memory (STM) activity pattern X across a field of feature detectors F\textsubscript{1}. The input I also excites an orienting subsystem A, but pattern X at F\textsubscript{1} inhibits A before it can generate an output signal. Activity pattern X also elicits an output pattern S which, via the bottom-up adaptive filter, instates an STM activity pattern Y across a category representation field, F\textsubscript{2}. In Figure 1b, pattern Y reads a top-down template pattern V into F\textsubscript{1}. Template V mismatches input I, thereby significantly inhibiting STM activity across F\textsubscript{1}. The amount by which activity in X is attenuated to generate X* depends upon how much of the input pattern I is encoded within the template pattern V.

When a mismatch attenuates STM activity across F\textsubscript{1}, the total size of the inhibitory signal from F\textsubscript{1} to A is also attenuated. If the attenuation is sufficiently great, inhibition from F\textsubscript{1} to A can no longer prevent the arousal source A from firing. Figure 1c depicts how disinhibition of A releases an arousal burst to F\textsubscript{2} which, equally, or nonspecifically, excites all the F\textsubscript{2} cells. The cell populations of F\textsubscript{2} react to such an arousal signal in a state-dependent fashion. In the special case that F\textsubscript{2} chooses a single population for STM storage, the arousal burst selectively inhibits, or resets, the active population in F\textsubscript{2}. This inhibition is long-lasting. One physiological design for F\textsubscript{2} processing which has these properties is a gated dipole field\textsuperscript{5,6}. A gated dipole field consists of opponent processing channels which are gated, or multiplied, by habituating chemical transmitters. A nonspecific arousal burst induces selective and enduring inhibition of active populations within a gated dipole field.

In Figure 1c, inhibition of Y leads to removal of the top-down template V, and thereby terminates the mismatch between I and V. Input pattern I can thus reinstate the original activity pattern X across F\textsubscript{1}, which again generates the output pattern S from F\textsubscript{1} and the input pattern T to F\textsubscript{2}. Due to the enduring inhibition at F\textsubscript{2}, the input pattern T can no longer activate the original pattern Y at F\textsubscript{2}. A new pattern Y* is thus generated at F\textsubscript{2} by I (Figure 1d).

The new activity pattern Y* reads-out a new top-down template pattern V*. If a mismatch again occurs at F\textsubscript{1}, the orienting subsystem is again engaged, thereby leading to another arousal-mediated reset of STM at F\textsubscript{2}. In this way, a rapid series of STM matching and reset events may occur. Such an STM matching and reset series controls the system's search of long term memory (LTM) by sequentially engaging the novelty-sensitive orienting subsystem. Although STM is reset sequentially in time via this mismatch-mediated, self-terminating LTM search process, the mechanisms which control the LTM search are all parallel network interactions, rather than serial algorithms. Such a parallel search scheme continuously adjusts itself to the system's evolving LTM codes. In general, the spatial configuration of LTM codes depends upon both the system's initial configuration and its unique learning history, and hence cannot be predicted \textit{a priori} by a pre-wired search algorithm. Instead, the mismatch-mediated engagement of the orienting subsystem realizes a self-adjusting search.

The mismatch-mediated search of LTM ends when an STM pattern across
F₂ reads-out a top-down template which matches I, to the degree of accuracy required by the level of attentional vigilance (equation (23)), or which has not yet undergone any prior learning. In the latter case, a new recognition category is then established as a bottom-up code and top-down template are learned.

**ATTENTIONAL GAIN CONTROL AND PATTERN MATCHING: THE 2/3 RULE**

The STM reset and search process described above makes a paradoxical demand upon the processing dynamics of F₁: the addition of new excitatory top-down signals in the pattern V to the bottom-up signals in the pattern I causes a decrease in overall F₁ activity (Figures 1a and 1b). This property is due to the attentional gain control mechanism, which is distinct from attentional priming by the top-down template V. While F₂ is active, the attentional priming mechanism delivers excitatory specific learned template patterns to F₁. Top-down attentional gain control has an inhibitory nonspecific unlearned effect on the sensitivity with which F₁ responds to the template pattern, as well as to other patterns received by F₁. The attentional gain control process enables F₁ to tell the difference between bottom-up and top-down signals. In Figure 1a, during bottom-up processing, a suprathreshold node in F₁ is one which receives both a specific input from the input pattern I and a nonspecific attentional gain control input. In Figure 1b, during the matching of simultaneous bottom-up and top-down patterns, attentional gain control signals to F₁ are inhibited by the top-down channel. Nodes of F₁ must then receive sufficiently large inputs from both the bottom-up and the top-down signal patterns to generate suprathreshold activities. Nodes which receive a bottom-up input or a top-down input, but not both, cannot become suprathreshold: mismatched inputs cannot generate suprathreshold activities. Attentional gain control thus leads to a matching process whereby the addition of top-down excitatory inputs to F₁ can lead to an overall decrease in F₁’s STM activity. Since, in each case, an F₁ node becomes active only if it receives large signals from two of the three input sources, this matching process is called the 2/3 Rule. Simple input environments exist in which code learning is unstable if the 2/3 Rule is violated⁴. Below are summarized the equations for the simplest ART network, which is called ART 1. Mathematical properties of ART 1 are also summarized.

**NETWORK EQUATIONS: INTERACTIONS BETWEEN SHORT TERM MEMORY AND LONG TERM MEMORY PATTERNS**

The STM equations for F₁ and F₂ and LTM equations for the bottom-up and top-down adaptive filters will now be described in dimensionless form, where the number of parameters is reduced to a minimum.

**A. STM Equations**

The STM activity $x_k$ of any node $v_k$ in F₁ or F₂ obeys a membrane equation of the form

$$
\epsilon \frac{d}{dt} x_k = -x_k + (1 - Ax_k) J^+_k - (B + Cx_k) J^-_k,
$$

(1)

where $J^+_k$ is the total excitatory input to $v_k$, $J^-_k$ is the total inhibitory input to $v_k$, and all the parameters are nonnegative.

Nodes in F₁ are denoted by $v_i$, where $i = 1, 2, \ldots, M$. Nodes in F₂ are
The excitatory input $\mathcal{I}_x \epsilon \sum \mathcal{I}_x A = \mathcal{I}_x$

The bottom-up input is the sum of the feedback signal $\mathcal{B}(x)$ from $a$ to $x$ and the bottom-up adaptive filter input $\mathcal{J}_x$ from $x$ to itself. The excitatory input in equation (2) is the sum of a positive $\mathcal{J}_a$.

$\mathcal{G} \mathcal{I}_x + 1 > 1 \mathcal{B} + 1 \mathcal{I}_x$

(8)

Therefore, the $\mathcal{J}_x$ node $\mathcal{N}_x$ equals 0. Thus by (7), $\mathcal{N}_a$ becomes active if $\mathcal{I}_x - \mathcal{A}_x \mathcal{C}_x > 0$. The output threshold of $\mathcal{I}_x$ is active when $\mathcal{I}_x = 1$ but remains inactive when $\mathcal{I}_x = 0$. The output threshold requires that $\mathcal{I}_x$ becomes active when $\mathcal{I}_x > 1$. Therefore, equation (2) requires that $\mathcal{I}_x$ become active. When $\mathcal{I}_x > 1$, excitatory thresholds are maintained at a tonic subthreshold level; that is, $\mathcal{I}_x = 1$ when $\mathcal{N}_x(0) = 0$.

(6)

Setting $\mathcal{I}_x = 1$,

(5)

otherwise $\mathcal{I}_x \mathcal{B}_x = \mathcal{C}_x$

where $\mathcal{B}_x$ is the excitatory input to node $\mathcal{N}_x$ from the top-down pathway from $a$. Each excitatory input is given a value in the top-down pathway from $a$ to $x$. Each $\mathcal{I}_x$ input to $\mathcal{A}_x$ is the sum of excitatory activity $\mathcal{I}_x$ from $x$ to itself and the tonic $\mathcal{J}_x$.

(4)

$$\mathcal{I}_x(\mathcal{I}_x \mathcal{C}_x + \mathcal{B}_x) - \mathcal{I}_x(\mathcal{I}_x \mathcal{C}_x + \mathcal{B}_x) + \mathcal{J}_x = \mathcal{J}_x$$

The excitatory input to the node $\mathcal{N}_x$ is the sum of the excitatory input $\mathcal{I}_x$ from $x$ to itself and the tonic $\mathcal{J}_x$.

(3)

$$\mathcal{J}_x(\mathcal{I}_x \mathcal{C}_x + \mathcal{B}_x) - \mathcal{J}_x(\mathcal{I}_x \mathcal{C}_x + \mathcal{B}_x) + \mathcal{J}_x = \mathcal{J}_x$$

(2)

$$\mathcal{J}_x(\mathcal{I}_x \mathcal{C}_x + \mathcal{B}_x) - \mathcal{J}_x(\mathcal{I}_x \mathcal{C}_x + \mathcal{B}_x) + \mathcal{J}_x = \mathcal{J}_x$$

$\mathcal{J}_x \epsilon \sum \mathcal{I}_x A = \mathcal{I}_x$

(1)

Thus, where $\mathcal{J}_a$ denotes by the adaptive filter.
where \( h(x_i) \) is the signal emitted by the \( F_1 \) node \( v_i \) and \( z_{ij} \) is the LTM trace in the pathway from \( v_i \) to \( v_j \). Thus

\[
J_j^+ = g(x_j) + T_j. \tag{10}
\]

Input \( J_j^- \) adds up negative feedback signals \( g(x_k) \) from all the other nodes in \( F_2 \):

\[
J_j^- = \sum_{k \neq j} g(x_k). \tag{11}
\]

Taken together, the positive feedback signal \( g(x_j) \) in (10) and the negative feedback signal \( J_j^- \) in (11) define an on-center off-surrond feedback interaction which contrast-enhances the STM activity pattern \( Y \) of \( F_2 \) in response to the input pattern \( T \).

The parameters of \( F_2 \) can be chosen so that this contrast-enhancement process enables \( F_2 \) to choose for STM activation only the node \( v_j \) which receives the largest input \( T_j \). Then when parameter \( \epsilon \) is small, \( F_2 \) behaves approximately like a binary switching, or choice, circuit:

\[
f(x_j) = \begin{cases} 1 & \text{if } T_j = \max\{T_k\} \\ 0 & \text{otherwise.} \end{cases} \tag{12}
\]

In the choice case, the top-down template in (4) obeys

\[
V_i = \begin{cases} D_1 z_{ji} & \text{if the } F_2 \text{ node } v_j \text{ is active} \\ 0 & \text{if } F_2 \text{ is inactive.} \end{cases} \tag{13}
\]

In the choice case, then, when \( I \) is active and the \( F_2 \) node \( v_j \) is active,

\[
\frac{dx_i}{dt} = -x_i + (1 - A_1 x_i)(I_i + D_1 z_{ji}) - (B_1 + C_1 x_i) \\
= (I_i + D_1 z_{ji} - B_1) - x_i(1 + A_1(I_i + D_1 z_{ji}) + C_1). \tag{14}
\]

In the dimensionless equations, \( 0 \leq z_{ij} \leq 1 \). The 2/3 Rule requires that \( v_i \) remain active when \( I_i = 1 \) and \( z_{ji} = 1 \), but become inactive when either \( I_i = 0 \) or \( z_{ji} = 0 \). By (14), \( x_i \) remains positive iff \( I_i + D_1 z_{ji} > B_1 \). Thus implementation of the 2/3 Rule when \( F_2 \) is active places constraint (15) on the strength of the patterned input signals:

\[
\max\{1, D_1\} < B_1 < 1 + D_1. \tag{15}
\]

The 2/3 Rule implies that if the top-down LTM trace \( z_{ji} \) becomes smaller than some critical value \( \bar{z} \), then when \( v_j \) is active, \( v_i \) will be inactive even if \( I_i = 1 \). That is, the feature represented by the \( F_1 \) node \( v_i \) will drop out of the critical feature pattern coded by \( v_j \). By (14) and (15),

\[
\bar{z} = \frac{B_1 - 1}{D_1}. \tag{16}
\]
B. LTM Equations

The LTM trace of the bottom-up pathway from $v_i$ to $v_j$ obeys a learning equation of the form

$$\frac{d}{dt}z_{ij} = K_1 f(x_j)[-E_{ij}z_{ij} + h(x_i)], \quad (17)$$

where

$$h(x_i) = \begin{cases} 1 & \text{if } x_i > 0 \\ 0 & \text{if } x_i \leq 0. \end{cases} \quad (18)$$

In (17), term $f(x_j)$ is a postsynaptic sampling, or learning, signal because $f(x_j) = 0$ implies $\frac{d}{dt}z_{ij} = 0$. Term $f(x_j)$ is also the output signal of $v_j$ to pathways from $v_j$ to $F_1$, as in (4).

The LTM trace of the top-down pathway from $v_j$ to $v_i$ also obeys a learning equation of the form

$$\frac{d}{dt}z_{ji} = K_2 f(x_j)[-E_{ji}z_{ji} + h(x_i)]. \quad (19)$$

In the present model, the simplest choice of $K_2$ and $E_{ji}$ was made for the top-down LTM traces:

$$K_2 = E_{ji} = 1. \quad (20)$$

A more complex choice of $E_{ij}$ was made for the bottom-up LTM traces in order to generate the Weber Law Rule, which is needed to achieve direct access to codes for arbitrary input environments after learning self-stabilizes. The Weber Law Rule requires that the positive bottom-up LTM traces learned during the encoding of an $F_1$ pattern $X$ with a smaller number $|X|$ of active nodes be larger than the LTM traces learned during the encoding of an $F_1$ pattern with a larger number of active nodes, other things being equal. This inverse relationship between pattern complexity and bottom-up LTM trace strength can be realized by allowing the bottom-up LTM traces at each node $v_j$ to compete among themselves for synaptic sites. The Weber Law Rule can also be generated by the STM dynamics of $F_1$ when competitive interactions are assumed to occur among the nodes of $F_1$.

Competition among the LTM traces which abut the node $v_j$ is modelled by defining

$$E_{ij} = h(x_i) + L^{-1} \sum_{k \neq i} h(x_k) \quad (21)$$

and letting $K_1 =$ constant. It is convenient to write $K_1$ in the form $K_1 = KL$. A physical interpretation of this choice can be seen by rewriting (17) in the form

$$\frac{d}{dt}z_{ij} = K f(x_j)[(1 - z_{ij})Lh(x_i) - z_{ij} \sum_{k \neq i} h(x_k)]. \quad (22)$$

By (22), when a postsynaptic signal $f(x_j)$ is positive, a positive presynaptic signal from the $F_1$ node $v_i$ can commit receptor sites to the LTM process $z_{ij}$ at a rate $(1 - z_{ij})Lh(x_i)K f(x_j)$. In other words, uncommitted sites—which number
(1 - z_{ij}) out of the total population size 1—are committed by the joint action of
signals Lh(x_i) and Kf(x_j). Simultaneously signals h(x_k), k ≠ i, which reach v_j
at different patches of the v_j membrane, compete for the sites which are already
committed to z_{ij} via the mass action competitive terms −z_{ij}h(x_k)Kf(x_j). In
other words, sites which are committed to z_{ij} lose their commitment at a rate
−z_{ij} \sum_{k ≠ i} h(x_k)Kf(x_j) which is proportional to the number of committed sites
z_{ij}, the total competitive input − \sum_{k ≠ i} h(x_k), and the postsynaptic gating signal
Kf(x_j).

C. STM Reset System

A simple type of mismatch-mediated activation of A and STM reset of F_2
by A were implemented for binary inputs. Each active input pathway sends
an excitatory signal of size P to the orienting subsystem A. Potentials x_i of F_1
which exceed zero generate an inhibitory signal of size Q to A. These constraints
lead to the following Reset Rule.

Population A generates a nonspecific reset wave to F_2 whenever

\[ \frac{|X|}{|I|} < \rho = \frac{P}{Q} \] (23)

where I is the current input pattern, |X| is the number of nodes across F_1 such
that x_i > 0, and \( \rho \) is called the vigilance parameter. The nonspecific reset wave
successively shuts off active F_2 nodes until the search ends or the input pattern
I shuts off. Thus (12) must be modified as follows to maintain inhibition of all
F_2 nodes which have been reset by A during the presentation of I:

\[ f(x_j) = \begin{cases} 1 & \text{if } T_j = \max \{T_k : k \in J \} \\ 0 & \text{otherwise} \end{cases} \] (24)

where J is the set of indices of F_2 nodes which have not yet been reset on the
present learning trial. At the beginning of each new learning trial, J is reset at
\{M + 1 \ldots N\}. As a learning trial proceeds, J loses one index at a time until
the mismatch-mediated search for F_2 nodes terminates.

THEOREMS WHICH CHARACTERIZE THE GLOBAL
DYNAMICS OF THE ART 1 SYSTEM

A series of theorems^4 analyze the global dynamics of the ART system. The
theorems are proved for the case that the input patterns are binary and that
"fast learning" occurs, i.e., that the LTM traces approach their equilibrium
values on each trial. With these hypotheses, the learning process is shown to
self-stabilize. That is, after a finite number of trials, the learned critical feature
pattern associated with each F_2 node remains constant. Thereafter, each input
directly accesses that category whose critical feature pattern matches it best.
This self-stabilization property does not require the assumption that plasticity
is turned off, i.e., that \( K_1 \) in (17) and \( K_2 \) in (19) approach 0 after some finite
interval. The length of time needed for the code to self-stabilize depends only
upon the complexity of the set of input patterns, and is not set externally or a
priori.

The theorems further specify details of system dynamics. For example, each
LTM strength \( z_{ij}(t) \) and \( z_{ji}(t) \) is shown to oscillate at most once as learning pro-
ceeds. This occurs despite the fact that, in a complex input environment, many
searches and category recodings may occur before the system self-stabilizes. Thus the learning process is remarkably stable. Also, given an arbitrary learning history, the order of search elicited by any input is characterized. The order of search is determined by bottom-up $F_2$ inputs $T_j$. Note, however, that the sum $T_j$ depends upon both the pattern of STM activity across $F_1$ and the strengths of all the bottom-up LTM traces $z_{ij}$. Fluctuations which occur in these STM and LTM values could, in principle, destabilize the system as follows. First, the initial choice of an $F_2$ node depends only upon the $F_1$ (STM) activity pattern generated by I and the system's prior learning (LTM) history (Figure 1a). However, once $F_2$ becomes active, read-out of its template alters $F_1$ activity (Figure 1b). This read-out can dramatically alter the distribution of $T_j$ values. However, the theorems guarantee that the original $F_2$ choice is confirmed by template read-out, so search proceeds as in Figure 1. Once search ends, however, learning alters both the pattern of $F_1$ STM activity, via changes in the top-down LTM traces, and the $F_2$ input function $T_j$, via the bottom-up LTM traces. The theorems also guarantee that the $F_2$ choice is confirmed by learning. In sum, $F_2$ reset can occur only via the orienting subsystem, which is activated by a mismatch between the input pattern and the critical feature pattern of an active $F_2$ node. While the order of search depends upon the entire coding history of the network, the decision to end the search depends upon the matching criterion as determined by the vigilance parameter $\rho$.

The size of $\rho$ determines how coarse the learned recognition code will be. A small value of $\rho$ leads to coarse recognition categories, whereas a large value of $\rho$ leads to fine recognition categories. Environmental disconfirmation can increase $\rho$, thereby enabling the network to learn finer distinctions than it previously could. Using such a scheme, an alphabet of 26 letters can be classified in no more than 3 learning trials, at any level of vigilance.

REFERENCES