

Generalization, Discrimination, and Multiple Categorization Using Adaptive Resonance Theory

Pierre Lavoie, *Member, IEEE*, Jean-François Crespo, and Yvon Savaria, *Member, IEEE*

Abstract—The internal competition between categories in the adaptive resonance theory (ART) neural model can be biased by replacing the original choice function by one that contains an attentional tuning parameter under external control. For the same input but different values of the attentional tuning parameter, the network can learn and recall different categories with different degrees of generality, thus permitting the coexistence of both general and specific categorizations of the same set of data. Any number of these categorizations can be learned within one and the same network by virtue of generalization and discrimination properties. A simple model in which the attentional tuning parameter and the vigilance parameter of ART are linked together is described. The self-stabilization property is shown to be preserved for an arbitrary sequence of analog inputs, and for arbitrary orderings of arbitrarily chosen vigilance levels.

Index Terms—Adaptive resonance theory (ART), categorization, choice function, discrimination, generalization, neural network, stability, unsupervised.

I. INTRODUCTION

THE faculty of an unsupervised neural network to learn intersecting categories, and to respond by more than one category to a familiar input, is examined. Such a faculty is necessary for learning and recalling categories exhibiting various degrees of generality, for example, and for the coexistence of general and specific categories on an equal footing. By definition, neural networks based on the premise that all categories are mutually exclusive do not have this faculty.

A simple illustration of the desired faculty is shown in Fig. 1. In this example, a hypothetical neural network has previously learned categories corresponding to “flowers” and “tulips.” It can be tuned by means of a parameter controlling whether the recalled category should be general, or rather specific, that is, correspond to a broad, or rather a narrow portion of the input space. In response to an input describing tulips, when the tuning parameter is set to “general,” the network chooses the flowers category; and when the tuning parameter is set to “specific,” the same network chooses the tulips category.

Let us call *categorization* the mapping that associates categories with each input in a set. If a neural network can output only one category per input, the mapping is a many-to-one relation, and all the categories are mutually exclusive.

Manuscript received April 1, 1998; revised January 4, 1999 and March 23, 1999.

P. Lavoie is with the Defense Research Establishment-Ottawa, Ottawa, Ont., K1A 0Z4 Canada.

J.-F. Crespo is with Nortel Networks, Nuns Island, Verdun, Quebec, H3E 1H6 Canada.

Y. Savaria is with the École Polytechnique de Montreal, St. Centre-Ville, Montreal, Quebec, H3C 3A7 Canada.

Publisher Item Identifier S 1045-9227(99)05971-8.

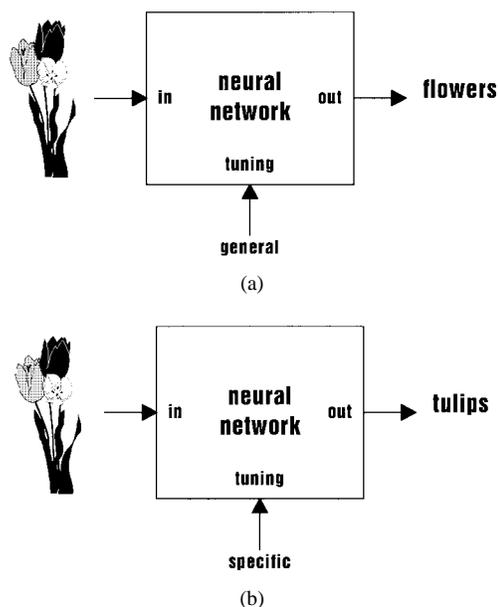


Fig. 1. A simple illustration of the multiple categorization faculty. A neural network has previously learned categories corresponding to “flowers” and “tulips.” It can recall either category in response to an input describing tulips. (a) Tuning parameter set to “general.” (b) Tuning parameter set to “specific.”

If a network can output multiple categories, the mapping can be a many-to-many relation, that is, categories can intersect and cover each other, and multiple categorization becomes possible. Alternatively, a neural network having a single output can be “tuned” to learn multiple categorizations and recall them in succession, as in Fig. 1.

In this paper, we examine multiple categorization in the context of adaptive resonance theory (ART), a theory of human cognitive information processing introduced by Grossberg, that addresses the stability-plasticity dilemma [1], [2]. More precisely, we use a self-organizing neural model called fuzzy ART by Carpenter *et al.* [3]. This neural network accepts analog inputs one at a time and develops a categorization. Familiar inputs activate their category, whereas unfamiliar inputs trigger either adaptive learning by an existing category, or commitment of a new category.

Fuzzy ART’s behavior lends itself well to simple geometrical interpretation owing to an internal representation of category prototypes as *hyperrectangles* in the input space. These hyperrectangles are allowed to overlap each other. Despite these overlaps, the developed categorization is a many-to-one relation and the categories are all mutually exclusive. This is due to the category choice process, by which fuzzy

ART always responds the same way to a familiar input: it recalls the smallest hyperrectangle containing this input [4]. Hyperrectangle overlaps have been argued to be an inconvenience if categories are mutually exclusive [5]. Our stance here is that these overlaps can be useful, and we describe a model that exploits them for multiple categorization.

In order to learn intersecting and overlapping categories, a neural network must be capable of repressing previously known categories while it forms new ones. In other words, it must be able to make temporary abstraction of previous knowledge. In the context of categories with various degrees of generality, this can be expressed as two complementary properties: if a neural network with previous knowledge of specific categories can learn new, broader ones, then we shall say that it has the *generalization* property; if it can do the reverse, then it has the *discrimination* property. Returning to the neural network of Fig. 1, generalization would allow the learning of the tulips category first, and the flowers category next, whereas discrimination would allow the reverse. In the case of fuzzy ART, increasing the value of a network parameter called *vigilance* allows formation of new, more specific categories intersecting broad ones that are already known. The network is thus capable of discrimination. However, reducing the same parameter value does not yield generalization. This is due to the predilection of fuzzy ART for the smallest hyperrectangle containing the input.

In this paper, it is shown that a simple modification grants to fuzzy ART both the generalization property and the ability to tune its attentional subsystem to categories of a desired degree of generality. What is remarkable is that no additional structural or functional unit is required: replacing the fixed “bottom-up adaptive filter” [6] by a tunable filter is sufficient to make the internal competition between the categories context-sensitive. A new tuning parameter need not be invented, as the readily available vigilance parameter can control both vigilance and attentional tuning. The resulting model is very simple and capable of multiple categorization for discrete and continuous degrees of generality. Returning to the example of Fig. 1, the modified fuzzy ART can learn the flowers and tulips categories in either order. Once these categories are learned, the network can choose either one in response to an input describing tulips, its decision being based on the value of the vigilance parameter. The self-stabilization property is preserved for arbitrary orderings of arbitrarily chosen vigilance levels. The model is applicable to tuning criteria other than generality, like shape, orientation, intensity, and so on. This approach is different from, and yet compatible with, models proposed for hierarchical categorization, a challenging problem where part/whole relationships are formed between categories having different levels of generality. A review of ART-based models for hierarchical clustering can be found in [7].

The fuzzy ART algorithm, input normalization, and order of search through the categories are reviewed in the next section. Since extensive literature exists on fuzzy ART and other ART models [3], [4], [6], [8]–[11], only the algorithmic aspects pertinent to multiple categorization are reviewed. In Section III, a “tunable” choice function is introduced, and the resulting order of search through the categories is compared to that

of the original choice function. In Section IV, learning with variable vigilance is examined and shown not to prevent self-stabilization of the weights for an arbitrary sequence of inputs. Computer simulation results are presented in Section V.

II. FUZZY ART

To avoid a category proliferation problem that could otherwise occur [12], Carpenter *et al.* recommend input normalization by a procedure called complement coding [3]. Let \mathbf{a} be an M -dimensional vector (a_1, a_2, \dots, a_M) , where $0 \leq a_i \leq 1$. The complement coded input \mathbf{I} is obtained as

$$\begin{aligned} \mathbf{I} &= (a_1, a_2, \dots, a_M, 1 - a_1, 1 - a_2, \dots, 1 - a_M) \\ &= (\mathbf{a}, \mathbf{a}^c). \end{aligned} \quad (1)$$

Assign to each category j a vector $\mathbf{w}_j = (w_{j1}, w_{j2}, \dots, w_{j2M})$ of adaptive weights. Each category is initially uncommitted, and its weights are initialized to one. The functionality of Fuzzy ART may be described as a three-step algorithm [3].

Step 1) Category Choice: Upon presentation of an input \mathbf{I} , a choice function T_j is computed for each category j

$$T_j = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|}. \quad (2)$$

The norm operator $|\cdot|$ is defined as $|\mathbf{x}| \equiv \sum_{i=1}^{2M} |x_i|$, the symbol \wedge denotes the fuzzy AND operator, that is, $\mathbf{x} \wedge \mathbf{y} \equiv (\min(x_1, y_1), \dots, \min(x_{2M}, y_{2M}))$, and α is a user-defined parameter, $\alpha > 0$. The category J for which the choice function is maximal, that is, $T_J = \max\{T_j, j = 1, 2, 3, \dots\}$, is chosen for the vigilance test.

Step 2) Vigilance Test: The similarity between \mathbf{w}_J and \mathbf{I} is compared to a parameter ρ called *vigilance*, $0 \leq \rho \leq 1$, in the following test:

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} \geq \rho. \quad (3)$$

If the test is passed, then resonance occurs [Step 3]) and learning takes place. If the test is failed, then mismatch reset occurs: the value of T_j is set to -1 for the duration of the current input presentation, another category is chosen in Step 1), and the vigilance test is repeated. Categories are searched, that is, chosen and then tested, until one that meets (3) is found. This category is said to be *selected* for \mathbf{I} . It is either already committed or uncommitted, in which case it becomes committed during resonance.

Step 3: Resonance: Resonance makes reference to the internal dynamics of the neural network as it pays attention to the vector $(\mathbf{I} \wedge \mathbf{w}_J)$. During resonance, the weight vector \mathbf{w}_J of the selected category is updated according to the equation

$$\mathbf{w}_J^{(\text{new})} = \beta(\mathbf{I} \wedge \mathbf{w}_J^{(\text{old})}) + (1 - \beta)\mathbf{w}_J^{(\text{old})} \quad (4)$$

where β is a learning rate parameter, $0 < \beta \leq 1$. What is learned is not the input \mathbf{I} itself, but rather an attended weight vector $(\mathbf{I} \wedge \mathbf{w}_J^{(\text{old})})$: fuzzy ART thus learns prototypes, rather than exemplars. The special case $\beta = 1$ is called fast learning and is assumed throughout this work. Once resonance is finished, a new input may be presented, and the three steps repeated.

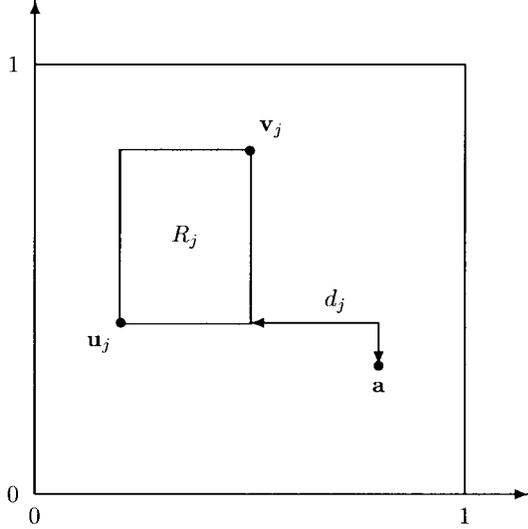


Fig. 2. Two-dimensional geometric interpretation of a category prototype $\mathbf{w}_j = (0.2, 0.4, 0.5, 0.2)$ as a rectangle R_j with corners $\mathbf{u}_j = (0.2, 0.4)$ and $\mathbf{v}_j = (0.5, 0.8)$. The norm of \mathbf{w}_j , and the size of R_j , are equal to 1.3, and 0.7, respectively. The input vector $\mathbf{a} = (0.8, 0.3)$ is shown lying outside R_j . The city-block distance, d_j , from R_j to \mathbf{a} is equal to 0.4.

Each weight vector \mathbf{w}_j may be written in the form

$$\mathbf{w}_j = (\mathbf{u}_j, \mathbf{v}_j^c) \quad (5)$$

where \mathbf{u}_j and \mathbf{v}_j are M -dimensional vectors corresponding to the two opposite corners of a hyperrectangle R_j . A two-dimensional example is shown in Fig. 2.

The *size* of R_j is defined by [3]

$$|R_j| \equiv \begin{cases} |\mathbf{v}_j - \mathbf{u}_j|, & \text{if } j \text{ is committed} \\ -M, & \text{if } j \text{ is uncommitted.} \end{cases} \quad (6)$$

For brevity, we refer to $|R_j|$ as the size of category j . This size is related to the norm of the weight vector $|\mathbf{w}_j|$ by

$$|R_j| = M - |\mathbf{w}_j|. \quad (7)$$

Note that a large category has a short weight vector and vice versa.

With fast learning, (4) reduces to

$$\mathbf{w}_J^{(\text{new})} = \mathbf{I} \wedge \mathbf{w}_J^{(\text{old})} \quad (8)$$

and the corners of R_j are updated by

$$\begin{aligned} \mathbf{u}_j^{(\text{new})} &= \mathbf{a} \wedge \mathbf{u}_j^{(\text{old})} \\ \mathbf{v}_j^{(\text{new})} &= \mathbf{a} \vee \mathbf{v}_j^{(\text{old})} \end{aligned} \quad (9)$$

where \vee denotes the fuzzy OR operator, that is, $\mathbf{x} \vee \mathbf{y} \equiv (\max(x_1, y_1), \dots, \max(x_M, y_M))$. When a committed category is selected, R_J expands to the minimum hyperrectangle containing both R_J and the input \mathbf{a} . If \mathbf{a} lies inside of R_J , then R_J is unchanged. Hence, once a category j is committed, the size of its hyperrectangle can only grow or remain the same. The maximum size allowed is determined by the vigilance test, which can be written in the form [3]

$$|R_J^{(\text{new})}| \leq M(1 - \rho). \quad (10)$$

Setting the vigilance ρ close to one yields only small hyper-

rectangles, whereas setting it close to zero allows both small and large sizes.

The order in which categories are searched is uniquely specified by the choice function (2). For this work it is convenient to rewrite it as follows:

$$\begin{aligned} T_j &= \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|} \\ &= \frac{M - |R_j| - (|\mathbf{w}_j| - |\mathbf{I} \wedge \mathbf{w}_j|)}{\alpha + M - |R_j|} \\ &= \frac{M - |R_j| - |\mathbf{w}_j - (\mathbf{I} \wedge \mathbf{w}_j)|}{\alpha + M - |R_j|} \\ &= \frac{M - |R_j| - d_j}{\alpha + M - |R_j|} \end{aligned} \quad (11)$$

where

$$d_j \equiv d(\mathbf{w}_j, \mathbf{I} \wedge \mathbf{w}_j) \quad (12)$$

denotes Kosko's fuzzy Hamming distance, that is, $d(\mathbf{x}, \mathbf{y}) \equiv \sum_{i=1}^{2M} |x_i - y_i|$ [13]. (Note that (12) defines distance between hyperrectangles, whereas in [4] distance is defined between points). If the input vector \mathbf{a} lies outside the hyperrectangle R_j , then d_j is equivalent to the city-block distance between \mathbf{a} and the closest edge or corner of R_j ; if \mathbf{a} lies inside R_j , then d_j takes the value zero; and if j is uncommitted, then d_j takes the value M . An example where \mathbf{a} lies outside R_j is shown in Fig. 2.

The distance between R_j and \mathbf{a} plays an important role in the fuzzy ART dynamics. Suppose $d_j = 0$. Then category j is said to be a *subset choice* for \mathbf{I} . When a subset choice is selected, its weight vector is unchanged during resonance since the prototype is nothing but the weight vector itself, that is, $\mathbf{I} \wedge \mathbf{w}_J = \mathbf{w}_J$; in this case, previous knowledge is conserved. Next suppose $d_j > 0$. Then when category j is selected, the weights necessarily change during resonance because the prototype learned, $(\mathbf{I} \wedge \mathbf{w}_J)$, is different from \mathbf{w}_J . It can easily be verified that the value of the choice function, T_j , is thereby increased, and so is the likelihood of choosing the same category if \mathbf{I} is presented again [3]. Uncommitted categories are a special case where $\mathbf{I} \wedge \mathbf{w}_j = \mathbf{I}$, $\mathbf{I} \neq \mathbf{w}_j$, $d_j = M$, and $T_j = T_u = M/(\alpha + 2M)$. When an uncommitted category is chosen, the vigilance criterion is necessarily met, thereby ending the search through the categories. Hence the value T_u defines a threshold below which no committed category choice for \mathbf{I} is searched. In other words, a committed category choice j for \mathbf{I} can be searched only if it satisfies

$$T_j \geq \frac{M}{\alpha + 2M}. \quad (13)$$

It is apparent from (11) that the order in which categories are searched is fixed and independent of the vigilance level ρ . Categories of identical size are searched in increasing order of their distance from \mathbf{a} , that is, closest first and farthest last, while categories that are equally distant from \mathbf{a} are searched in increasing order of size, that is, smallest first and largest last. In the particular case where the weights are not changing, $d_j = 0$ for all the selected categories. As pointed out in [4], Fuzzy ART then systematically chooses the smallest subset choice for the current input.

III. TUNABLE CATEGORY RECALL

A. A Tunable Choice Function

After weights have stabilized over an entire input set, the choice function (2) ensures that the selected category is always the smallest subset choice for the current input. In this section, we explore the selection of other subset choices by replacing (2) with a choice function that is tunable. Note that this choice function is different from those examined by Carpenter and Gjaja [14], and by Blume *et al.* [15].

Let us denote by \mathcal{S} a parameter equal to the desired category size $0 \leq \mathcal{S} \leq M$. This parameter specifies whether the recalled category should be general, or rather specific, that is, correspond to a large, or rather a small hyperrectangle. In order to tune the attentional subsystem of fuzzy ART to categories of size close to \mathcal{S} , we seek a choice function that fulfills many requirements, such as the following.

- 1) Subset choices must be searched in increasing order of the absolute difference between their size and \mathcal{S} .
- 2) Categories at a distance d from the input must be searched in increasing order of the absolute difference between their size and $(\mathcal{S} - d)$.
- 3) Categories of identical size $|R|$ must be searched in increasing order of their distance from the input, that is, closest first and farthest last.
- 4) Subset choices of the desired size \mathcal{S} must meet the vigilance criterion.
- 5) Learning must increase the value of the choice function for the selected category.

The first two requirements define the tuning property, the next two stem from the fuzzy ART dynamics, and the last one is essential for stability, as will be shown in Section IV. The choice function (2) meets all five requirements when $\mathcal{S} \equiv 0$, that is, when the desired size is fixed and equal to zero.

One tunable choice function that meets the above requirements is convex with its maximum at size \mathcal{S} . Changing the value of the tuning parameter translates the maximum with respect to category size, allowing activation of more than one category per input. This choice function is best introduced geometrically. Let us write a first term,

$$(M - d_j)^2$$

to promote categories with hyperrectangles in the vicinity of input vector \mathbf{a} . This term is independent of category size. Next, let us write a second term,

$$-(|R_j| - \mathcal{S} + d_j)^2$$

to promote categories having a postresonance size close to \mathcal{S} , that is, a preresonance size close to $(\mathcal{S} - d_j)$. It is a quadratic expression with respect to size $|R_j|$, reaching a maximum of zero at $|R_j| = \mathcal{S} - d_j$. Adding these closeness and sizing terms together yields

$$T_j = (M - d_j)^2 - (|R_j| - \mathcal{S} + d_j)^2. \quad (14)$$

The tunable choice function (14) meets all five aforementioned requirements, provided that $0 \leq \mathcal{S} \leq M(1 - \rho)$, as is shown in Appendix A.

The choice function (14) may be rewritten in conventional terms

$$\begin{aligned} T_j &= (M - d_j)^2 - (|R_j| - \mathcal{S} + d_j)^2 \\ &= M^2 - 2d_jM + d_j^2 - (|R_j| - \mathcal{S})^2 \\ &\quad - 2d_j(|R_j| - \mathcal{S}) - d_j^2 \\ &= (-|R_j| + \mathcal{S} + M - 2d_j)(|R_j| - \mathcal{S} + M) \\ &= (2(M - |R_j| - d_j) - (M - |R_j|) + \mathcal{S}) \\ &\quad \cdot (2M - (M - |R_j|) - \mathcal{S}) \\ &= (2|\mathbf{I} \wedge \mathbf{w}_j| - |\mathbf{w}_j| + \mathcal{S})(|\mathbf{w}_j^c| - \mathcal{S}). \end{aligned} \quad (15)$$

Thus T_j is equal to the product of two factors: one is a function of a similarity term, $|\mathbf{I} \wedge \mathbf{w}_j|$, and the norm of the weight vector, $|\mathbf{w}_j|$; the other is a function of the norm of the complement of the weight vector, $|\mathbf{w}_j^c|$.

The desired size \mathcal{S} can conveniently be defined as a function of ρ , the vigilance parameter, as long as the constraint $\mathcal{S} \leq M(1 - \rho)$ is satisfied. In particular, one may set

$$\mathcal{S} \equiv \kappa M(1 - \rho) \quad (16)$$

where κ is a user-defined constant such that $0 \leq \kappa \leq 1$. Then a single parameter controls both vigilance and attentional tuning.

Using (14), categories of identical size are searched closest first and farthest last, like with the choice function (12). However, categories at a distance d from the input \mathbf{a} are now searched in accordance with an inverted parabola that reaches a maximum of $(M - d)^2$ at size $(\mathcal{S} - d)$. If $\kappa = 1$, then the peak of the parabola corresponds to the largest size that can meet the vigilance criterion, $|R_j| \leq M(1 - \rho) - d_j$. In this case, categories that meet the criterion are searched in decreasing order of size, that is, largest first and smallest last, rather than in increasing order of size as with the choice function (2). Moreover, whenever the weights are not changing, the selected category becomes the largest subset choice whose size does not exceed $M(1 - \rho)$, rather than the smallest subset choice irrespective of ρ .

If the vigilance level is changed, for example from $\rho = 0.85$ to $\rho = 0.65$, then the choice function reaches its maximum at $|R_j|/M = 0.35$ instead of $|R_j|/M = 0.15$, and the attentional subsystem of fuzzy ART becomes tuned to larger categories. Therefore changing ρ allows more than one category to be activated for a same input. In fact, it can be shown that every committed category can be recalled for some prior input if ρ is properly set.

B. Stable Category Learning with Fixed Vigilance

An equilibrium state is considered to be reached when, upon presentation of a set of inputs, the network weights are not changing [16]. This is equivalent to saying that learning has stabilized. Generally, there can be many possible equilibrium states for an input set; which one is reached depends on the order of presentation of the inputs.

Of theoretical significance is a definition of stability that requires reaching an equilibrium state after just a single presentation of the input set, that is, a single learning trial. Carpenter *et al.* have shown that the original fuzzy ART

algorithm is stable in this sense if and only if the inputs are normalized, $\beta = 1$ (fast learning), and $\alpha \rightarrow 0$ [3]. The limit $\alpha \rightarrow 0$ is called the *conservative limit*, and sets fuzzy ART in its most stable operating state.

The astute reader may have noticed that, unlike the original choice function (2), the choice function (14) does not contain a choice parameter α . This parameter was omitted earlier for the sake of clarity. But without it, a conservative limit cannot be attained. A tunable choice function with a choice parameter is given by

$$T_j = (M - d_j)^2 - \frac{\alpha M}{M - S(1 - \alpha)} (|R_j| - S + d_j)^2 \quad (17)$$

where

$$S \equiv \kappa M(1 - \rho) \quad (18)$$

and

$$.3611 < \kappa \leq 1. \quad (19)$$

As shown in a lemma in Appendix B, the lower bound on κ guarantees that all the subset choices meeting the vigilance criterion remain above the threshold defined by uncommitted categories, even when α is high and ρ is low.

Using (17), the smaller α is, the flatter the parabolas are. At the conservative limit, the second term vanishes, placing subset choices on a straight horizontal line above all other choices. The larger α is, the narrower the subset choice parabola is, yielding a search that is more selective in terms of category size. Note that in the particular case where $\alpha = 1$, (17) simply reduces to (14).

Like in (2), choosing α close to zero predisposes fuzzy ART to stability rather than plasticity, and the limit $\alpha \rightarrow 0$ guarantees that the subset choices for an input \mathbf{I} take precedence over all other categories.

Lemma 1 (Learning on a Single Trial): Assume fuzzy ART with the tunable choice function (17), $\beta = 1, 0 \leq \rho \leq 1$, and $\alpha \rightarrow 0$. An equilibrium state is reached after a single presentation of a list of arbitrarily chosen analog inputs.

A proof is given in Appendix C.

IV. MULTIPLE CATEGORIZATION

A. Discrimination and Generalization

In the tunable choice function (14), ρ acts as a dual-purpose vigilance-tuning parameter. It simultaneously limits the category growth, and tunes the attentional subsystem to categories of a certain size. To tune the network, ρ must be variable, that is, free to change between input presentations, without resetting the network weights. (The idea of varying vigilance is not new. For instance, empirical results on the use of an adaptive vigilance strategy in the vigilant net are presented by Burke [17].) In this section we analyze the effect of a variable vigilance on the category learning process and on the self-stabilization of fuzzy ART. Self-stabilization with fixed vigilance has been documented at length [3], [6], [11].

Suppose broad categories are formed by a fuzzy ART network during a learning trial held at a low vigilance level

ρ^- . If another trial is held at a higher vigilance level ρ^+ using as the initial weights those obtained at the low vigilance, then more specific categories will be learned. Hence, increasing the value of ρ allows formation of new smaller categories inside of large ones that are already known. This property can be stated in a lemma.

Lemma 2 (Discrimination): After fuzzy ART with the choice function (2) or (17) has reached an equilibrium state at ρ^- , it may learn more at a higher vigilance level ρ^+ , $\rho^- < \rho^+$, from the same inputs.

Proof: By (10) subset choices activated at ρ^- may fail the vigilance test at ρ^+ . Uncommitted categories and nonsubset choices may thus be selected and trigger additional learning. ■

Decreasing the value of ρ should yield generalization, that is, the formation of broader categories in addition to specific ones that already exist. The next lemma, whose proof is given in Appendix D, states that fuzzy ART is not capable of generalization with the choice function (2), but that it is with (17).

Lemma 3 (Generalization): Assume fuzzy ART has reached an equilibrium state at ρ^+ . With the choice function (2), it remains in the same equilibrium state as long as $\rho \leq \rho^+$, whereas with the choice function (17), it can learn more from the same inputs when $\rho < \rho^+$.

The tunable choice function (17) allows fuzzy ART to learn multiple categorizations as ρ is increased or decreased, and to access each categorization through its own vigilance level. Returning to the example of Fig. 1, it can thus learn the tulips and flowers categories in either order, and recall either category in response to an input describing tulips.

B. Stable Category Learning with Variable Vigilance

It is imperative to determine whether the proposed model self-stabilizes across several vigilance levels, and if so, whether the vigilance levels can be arbitrarily chosen. Let us first examine the effect of alternating vigilance between two levels, ρ^- and ρ^+ . After learning has stabilized once at ρ^- and then once at ρ^+ , $\rho^- < \rho^+$, additional learning may still take place upon returning to ρ^- . Indeed categories that were previously undersized at ρ^- may be selected during the trial held at ρ^+ . The hyperrectangles of these categories may grow, making them more attractive for selection and further growth back at ρ^- . (By contrast, the original fuzzy ART would remain in equilibrium from ρ^+ to ρ^- by lack of generalization.) This leads to a lemma, whose proof is given in Appendix E.

Lemma 4: Assume fuzzy ART with the choice function (17), and let $0 \leq \rho^- < \rho^+ \leq 1$. By holding trials alternately at ρ^- and at ρ^+ , the number of categories satisfying

$$M(1 - \rho^+) \leq |R_j| \leq M(1 - \rho^-) \quad (20)$$

can increase or remain the same, but not decrease.

Now let us define *self-stabilization* as reaching an equilibrium state on a finite number of learning trials for any finite set of inputs. This prevents the creation of an infinite number of categories, or perpetual category updates, upon repeated

presentations of some inputs. This definition has been used for analyzing fuzzy ART's stability for a fixed vigilance [3], [11], [12], [18].

According to this definition, fuzzy ART is stable for a variable vigilance, provided the choice function meets the fifth requirement of Section III (a mild condition). This can be stated in a theorem, whose proof is given in Appendix F.

Theorem 1 (Stable Multiple Categorization): Assume fuzzy ART with $\beta = 1$ (fast learning), and a choice function such that learning necessarily increases the choice function value of the selected category. An equilibrium state is reached on a finite number of learning trials, in response to arbitrary orderings of arbitrarily chosen analog inputs, when ρ varies from trial to trial with values arbitrarily taken from a finite set $0 \leq \rho \leq 1$.

Corollary 1: Fuzzy ART with the choice function (17) has the stable multiple categorization property for any value of the parameter α , $\alpha > 0$.

Proof: By virtue of (17), learning increases the value of the choice function of the selected category, that is, if $d_J \neq 0$, then $T_J^{(\text{new})} > T_J^{(\text{old})}$. Indeed

$$\begin{aligned} T_J^{(\text{new})} - T_J^{(\text{old})} &= M^2 - \frac{\alpha M}{M - \mathcal{S}(1 - \alpha)} (|R_J^{(\text{new})}| - \mathcal{S})^2 \\ &\quad - (M - d_J)^2 \\ &\quad + \frac{\alpha M}{M - \mathcal{S}(1 - \alpha)} (|R_J^{(\text{old})}| - \mathcal{S} + d_J)^2 \\ &= M^2 - (M - d_J)^2 \\ &> 0. \end{aligned} \quad (21)$$

The stable multiple categorization property implies that there exists no finite list of inputs and vigilance levels that could trigger the creation of an infinite number of categories, or produce endless category updates. This self-stabilization is essential in many critical applications. Vigilance can thus be increased or decreased in any number of steps, which can be of any size. It can also be chosen at random before each learning trial or each input presentation. Vigilance levels that are close together will share many categories, whereas vigilance levels that are far apart will share few, if any, categories.

C. Abstraction Criterion

In some applications it may be useful to reinforce generalization by completing the vigilance test with an *abstraction test*. Indeed the generalization provided by the choice function (17) is "gentle" in comparison with the discrimination embodied in the vigilance test: small subset choices can always be selected for lack of better nonsubset choices, whereas the vigilance test sets a hard limit to category size even if this implies creating new categories. The object of an abstraction test would mirror that of the vigilance test, and aim at rejecting categories that are small and close to the input, while accepting ones that are larger and more distant. One possible abstraction test is given by

$$|R_J^{(\text{new})}| \geq \kappa(1 - \kappa)M(1 - \rho) \quad (22)$$

or

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} \leq 1 - \kappa(1 - \kappa)(1 - \rho) \quad (23)$$

where κ now denotes an *abstraction* parameter, $0.3611 < \kappa \leq 1$. Setting κ close to 0.3611 yields large hyperrectangles only, whereas setting it close to one allows both small and large sizes. It is imperative for stability that newly committed categories pass the abstraction test. This can be achieved by modifying the learning mechanism so that categories are committed with a minimum size equal to $\kappa(1 - \kappa)M(1 - \rho)$.

V. SIMULATION RESULTS

A. Flowers Data

First we illustrate the multiple categorization faculty by pursuing the example of Fig. 1. Two-dimensional fictitious data belonging to three classes labeled "tulips," "roses," and "sticks," are presented to a fuzzy ART neural network with the choice function (14), while the vigilance level is varied between 0.9, 0.75, and 0.5. Labels are of course withheld from the network, and the network weights are not reset on changes of the vigilance level.

A geometrical interpretation of the results of this simulation is shown in Fig. 3. Once learning has stabilized across all the vigilance levels, a total of five categories can be recalled by the network: three correspond to tulips, roses and sticks, a fourth corresponds to "flowers," that is, a combination of tulips and roses, and a fifth corresponds to all the inputs. The category selected for a given input depends on the current vigilance level. For example, in response to an input describing tulips, when $\rho = 0.9$, the network selects the tulips category; when $\rho = 0.75$, the same network selects the flowers category; and when $\rho = 0.5$, it selects the category in which all the patterns are lumped together. The hyperrectangles selected when the vigilance is equal to 0.9, 0.75, and 0.5 are shown in Fig. 3(a), (b), and (c), respectively.

B. Binary Patterns

Fig. 4 shows the results of a computer simulation in which binary patterns from [6] are presented to two fuzzy ART neural networks. The left part of the figure shows four patterns labeled A, B, C, and D being repeatedly presented as time elapses from the top down. The vigilance level used is shown next to each pattern. From trial to trial, low vigilance ($\rho = 0.5$) alternates with high vigilance ($\rho = 0.8$). The middle part of the figure shows the categories learned and recalled by a fuzzy ART neural network using the original choice function (2) with $\alpha = 0.1M$. A number appearing under a category indicates that this category is searched, and the letters "RES" indicate resonance. Last, the right part of the figure shows the categories learned and recalled when the tunable choice function (14) is used. Throughout the simulation both networks are set to fast learning ($\beta = 1$).

In this example, the two networks learn the same categories. After the first trial, they have learned Category 1, which corresponds to {A,B,C,D}. After the second trial, they have

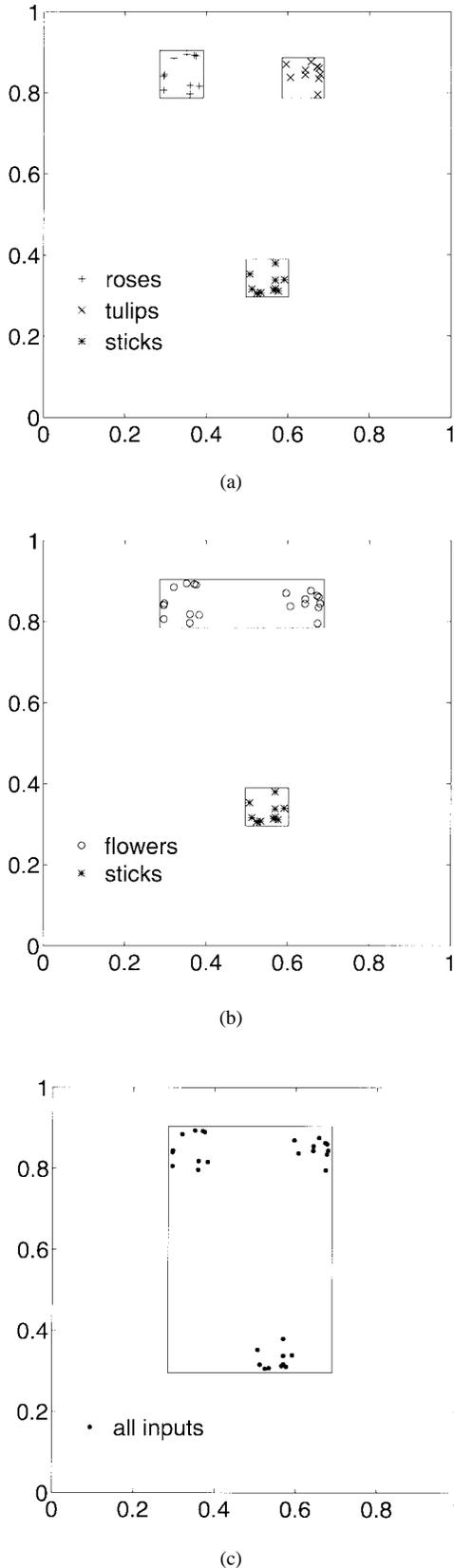


Fig. 3. A fuzzy ART neural network modified with the choice function (14) has learned five categories while the vigilance level was varied between 0.9, 0.75, and 0.5 without resetting the network weights. After stabilization, the categories recalled are shown for the three vigilance levels, in (a), (b), and (c), respectively. Rectangles and category labels are provided for convenience only. (a) $\rho = 0.9$ (b) $\rho = 0.75$ (c) $\rho = 0.5$

learned Categories 2 and 3, which correspond to $\{A,B\}$ and $\{C,D\}$, respectively.

Having learned the same three categories, the two networks recall them differently. In the case of the original choice function, whether $\rho = 0.5$ or $\rho = 0.8$, the same categories are recalled, namely Categories 2 and 3. Category 1 does not win anymore, even at low vigilance ($\rho = 0.5$). It has been eclipsed by the other two.

By contrast, the categories recalled by the tunable network change according to vigilance. When $\rho = 0.8$, Categories 2 and 3 are recalled, whereas when $\rho = 0.5$, Category 1 is recalled. Two distinct categorizations are thus activated by a change of vigilance.

VI. CONCLUSION

The choice function of fuzzy ART has been modified to include a tuning parameter under external control that can bias the internal competition between categories in favor of categories of a desired size. This attentional tuning mechanism allows recalling for a same input different categories under different circumstances, even when no additional learning takes place. This behavior cannot be duplicated by the conventional choice function, in which the order of search through the categories is fixed.

The modified choice function does not interfere with the orienting subsystem, the vigilance test, nor the learning of the fuzzy ART model. A single parameter ρ can be used both for vigilance and for attentional tuning. The resulting network can generalize and discriminate, and has been shown to reach an equilibrium state after a finite number of learning trials for arbitrary orderings of arbitrarily chosen values of the vigilance parameter. This work illustrates the flexibility of the ART framework, and its potential for a wide range of applications.

APPENDIX A

In this appendix, we present proofs that the choice function (14) meets the five requirements of Section III.

Lemma A.1: If $d_j = d_i = 0$ and $\|R_j| - S| < \|R_i| - S|$, then $T_j > T_i$.

Proof: If $d_j = d_i = 0$ and $\|R_j| - S| < \|R_i| - S|$, then

$$\begin{aligned} (|R_j| - S)^2 &< (|R_i| - S)^2 \\ M^2 - (|R_j| - S)^2 &> M^2 - (|R_i| - S)^2 \\ T_j &> T_i \end{aligned} \quad (24)$$

for all $i, j, i \neq j$. ■

Lemma A.2: If $d_j = d_i = d \neq 0$ and $\|R_j| - (S - d)| < \|R_i| - (S - d)|$, then $T_j > T_i$.

Proof: If $d_j = d_i = d$ and $\|R_j| - (S - d)| < \|R_i| - (S - d)|$, then

$$(M - d)^2 - (|R_j| - (S - d))^2 > (M - d)^2 - (|R_i| - (S - d))^2 \quad (25)$$

and $T_j > T_i$ for all $i, j, i \neq j$. ■

Lemma A.3: If $|R_j| = |R_i|$ and $d_j < d_i$, then $T_j > T_i$.

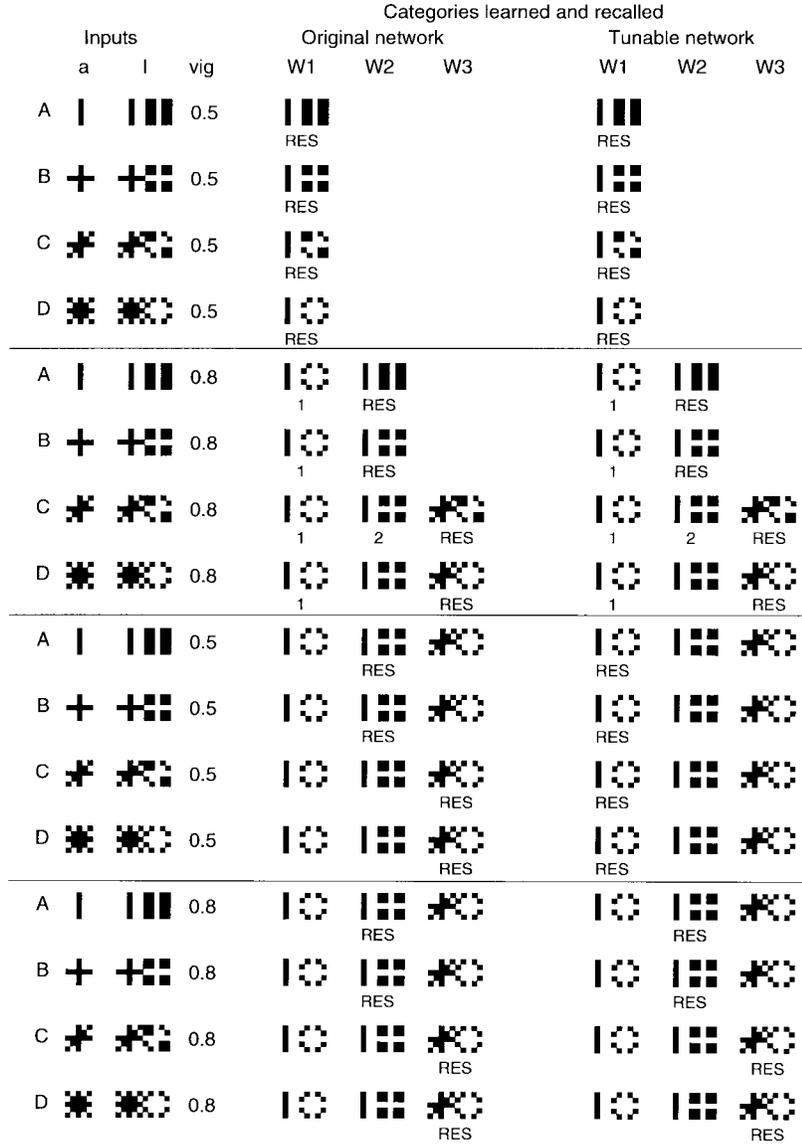


Fig. 4. Binary pattern learning example, showing the categories learned by fuzzy ART neural networks with the original and tunable choice functions side by side. Time elapses from the top down. The two networks simultaneously learn the same three categories, but recall them differently afterwards.

Proof: This pertains only to committed categories. (Indeed if a single one of i or j is uncommitted, then $|R_j| \neq |R_i|$, whereas if both i and j are uncommitted, then $d_j = d_i$.) The partial derivative of (14) with respect to distance is

$$\begin{aligned} \frac{\partial T_j}{\partial d_j} &= -2(M - d_j) - 2(|R_j| - S + d_j) \\ &= -2(M + |R_j| - S). \end{aligned} \quad (26)$$

Since $0 \leq S \leq M$, and for committed categories $0 \leq |R_j| \leq M$, we obtain

$$\frac{\partial T_j}{\partial d_j} < 0, \quad \text{if } j \text{ is committed.} \quad (27)$$

Lemma A.4: If $|R_j| = S$ and $d_j = 0$, then (10) is met.

Proof: If $d_j = 0$, then $|R_j^{(\text{new})}| = |R_j^{(\text{old})}| = S$, and by virtue of (10), the choice function (14) meets the requirement

if the desired size satisfies

$$S \leq M(1 - \rho). \quad (28)$$

Lemma A.5: If $d_j \equiv d_j^{(\text{old})} > 0$, then $T_j^{(\text{new})} > T_j^{(\text{old})}$. ■

Proof: If J is not a subset choice, then $0 < d_j \leq M$, and it can be verified that

$$\begin{aligned} T_j^{(\text{new})} - T_j^{(\text{old})} &= M^2 - (|R_j^{(\text{new})}| - S)^2 - (M - d_j)^2 \\ &\quad + (|R_j^{(\text{old})}| - S + d_j)^2 \\ &= M^2 - ((|R_j^{(\text{old})}| + d_j) - S)^2 \\ &\quad - (M - d_j)^2 + (|R_j^{(\text{old})}| - S + d_j)^2 \\ &= M^2 - (M - d_j)^2 \\ &> 0 \end{aligned} \quad (29)$$

holds true. ■

APPENDIX B

A lower-bound on the user-defined constant κ is derived.

Lemma B.1: If $0.3611 < \kappa \leq 1$, and $d_j = 0$, then $T_j \geq T_u$ for $T_u = -\mathcal{S}^2$, $\mathcal{S} \equiv \kappa M(1 - \rho)$, $0 \leq \rho \leq 1$, $0 \leq |R_j| \leq M(1 - \rho)$, and $\alpha > 0$.

Proof: Since the subset choice parabola reaches its minimum at size $|R_j| = M(1 - \rho) = \mathcal{S}/\kappa$, we must find κ that satisfies $T_j \geq T_u$, that is

$$\begin{aligned} T_j + \mathcal{S}^2 &= M^2 - \frac{\alpha M}{M - \mathcal{S}(1 - \alpha)} \left(\frac{\mathcal{S}}{\kappa} - \mathcal{S} \right)^2 + \mathcal{S}^2 \\ &= M^2 - \frac{\alpha M}{M - \mathcal{S}(1 - \alpha)} \mathcal{S}^2 \left(\frac{1 - \kappa}{\kappa} \right)^2 + \mathcal{S}^2 \\ &\geq 0 \end{aligned} \quad (30)$$

and thus

$$\left(\frac{1 - \kappa}{\kappa} \right)^2 \leq \frac{(M^2 + \mathcal{S}^2)(M - \mathcal{S} + \alpha \mathcal{S})}{\alpha M \mathcal{S}^2}. \quad (31)$$

To minimize the right hand term, we take the partial derivative with respect to α

$$\frac{\partial}{\partial \alpha} \left(\frac{(M^2 + \mathcal{S}^2)(M - \mathcal{S} + \alpha \mathcal{S})}{\alpha M \mathcal{S}^2} \right) = \frac{(M^2 + \mathcal{S}^2)(\mathcal{S} - M)}{\alpha^2 M \mathcal{S}^2} \quad (32)$$

and find that it is monotonically decreasing if $\mathcal{S} < M$, which is the case. At the limit $\alpha \rightarrow \infty$, (31) becomes

$$\left(\frac{1 - \kappa}{\kappa} \right)^2 \leq \frac{M^2 + \mathcal{S}^2}{M \mathcal{S}}. \quad (33)$$

Now taking the partial derivative of the right hand term with respect to \mathcal{S}

$$\frac{\partial}{\partial \mathcal{S}} \left(\frac{M^2 + \mathcal{S}^2}{M \mathcal{S}} \right) = \frac{\mathcal{S}^2 - M^2}{M \mathcal{S}^2} \quad (34)$$

we find that it is monotonically decreasing if $|\mathcal{S}| < M$, which is again the case. To maximize \mathcal{S} , we set $\rho = 0$, and (33) reduces to

$$\left(\frac{1 - \kappa}{\kappa} \right)^2 \leq \frac{1 + \kappa^2}{\kappa} \quad (35)$$

which can be written

$$\kappa^3 - \kappa^2 + 3\kappa - 1 \geq 0. \quad (36)$$

The real root of this third-degree polynomial is equal to 0.3611, and corresponds to the minimum value of κ guaranteeing that all the subset choices that can pass the vigilance test remain above threshold. ■

APPENDIX C

We prove Lemma 1, which states that under certain conditions, fuzzy ART with the tunable choice function (17) reaches an equilibrium state after a single presentation of an arbitrary list of analog inputs. The proof is based on another lemma and a corollary.

Lemma C.1: Assume fuzzy ART with the tunable choice function (17), $\beta = 1$, $0.3611 < \kappa \leq 1$, and fixed ρ . Once a category has been assigned to an input, it can always be searched, and pass the vigilance test for this input. This holds true for any value of the parameter α , $\alpha > 0$.

Proof: First, by (8) and the subset choice definition, once a category has coded an input \mathbf{I} , this category is and remains a subset choice for \mathbf{I} . Second, as a consequence of submitting all hyperrectangle growths to the vigilance test (10), subset choices continue to pass this test as long as ρ is held constant. Third, by Lemma B.1, the choice function of subset choices remains above the threshold T_u corresponding to uncommitted categories. ■

Corollary C.1: Assume the hypotheses of Lemma C.1. The number of categories committed is less than, or equal to the number of distinct inputs.

Proof: After one learning trial there is at least one subset choice for each input. By Lemma C.1 this subset choice can be searched ($T_j \geq T_u$), and can pass the vigilance test if the input is presented again. Therefore no uncommitted category can be searched, nor committed. ■

After one learning trial, there is at least one subset choice for every input. By Lemma C.1, all subset choices can be searched and pass the vigilance test, and at the conservative limit, (17) guarantees that these subset choices take precedence over all other categories. Consequently, in subsequent trials, all inputs activate subset choices, and no additional learning occurs. An equilibrium state is thus reached after a single trial. Note, however, that the final category assignment may differ from that obtained during this first trial.

APPENDIX D

Lemma 3 (Generalization): Assume fuzzy ART has reached an equilibrium state at ρ^+ . With the choice function (2), it remains in the same equilibrium state as long as $\rho \leq \rho^+$, whereas with the choice function (17), it can learn more from the same inputs when $\rho < \rho^+$.

Proof: First we prove that with (2) the categorization is fixed. Assume an equilibrium state at ρ^+ . When $\rho < \rho^+$, subset choices activated at ρ^+ necessarily pass the vigilance test and therefore no new category can be committed. Now suppose that when $\rho < \rho^+$, \mathbf{I} could activate a category J different from that activated at ρ^+ , say j , and that learning occurs, meaning $d_J \neq 0$. This implies $T_J \geq T_j$. But since the choice function is independent of ρ , it is J , instead of j , that would have been chosen for \mathbf{I} at ρ^+ . The only alternative is if J had failed the vigilance test at $\rho = \rho^+$, which would imply

$$\begin{aligned} |R_J| + d_J &> M(1 - \rho^+) \\ &\geq |R_j|. \end{aligned} \quad (37)$$

But then

$$\begin{aligned} T_J - T_j &= \frac{M - |R_J| - d_J}{\alpha + M - |R_J|} - \frac{M - |R_j|}{\alpha + M - |R_j|} \\ &= \frac{(M - |R_J| - d_J)(\alpha + M - |R_j|)}{(\alpha + M - |R_J|)(\alpha + M - |R_j|)} \\ &\quad - \frac{(M - |R_j|)(\alpha + M - |R_J|)}{(\alpha + M - |R_J|)(\alpha + M - |R_j|)} \\ &= \frac{\alpha(|R_j| - (|R_J| + d_J)) - d_J(M - |R_j|)}{(\alpha + M - |R_J|)(\alpha + M - |R_j|)} \\ &< 0 \end{aligned} \quad (38)$$

and j , instead of J , would be chosen for \mathbf{I} at $\rho < \rho^+$, contradicting our supposition. Therefore for all $\rho \leq \rho^+$, \mathbf{I} activates j , and the network remains in the same equilibrium state.

An example will suffice to prove that with (17) the categorization is mutable. Assume that $M = 1, \kappa = 1, \alpha = 1$, and $\rho^+ = 0.9$. In response to $\mathbf{I}_1 = (0.5, 0.5)$, a first category is committed, with resulting weight vector $\mathbf{w}_1 = (0.5, 0.5)$. In response to $\mathbf{I}_2 = (0.4, 0.6)$, Category 1 is chosen, passes the vigilance test

$$\frac{|\mathbf{I}_2 \wedge \mathbf{w}_1|}{M} = 0.9 \geq 0.9$$

and its weight vector becomes $\mathbf{w}_1 = (0.4, 0.5)$. In response to $\mathbf{I}_3 = (0.39, 0.61)$, Category 1 is chosen, and fails the vigilance test

$$\frac{|\mathbf{I}_3 \wedge \mathbf{w}_1|}{M} = 0.89 \not\geq 0.9.$$

A second category is committed, with resulting weight vector $\mathbf{w}_2 = (0.39, 0.61)$. At this point learning has stabilized at ρ^+ , and the inputs $\mathbf{I}_1, \mathbf{I}_2$, and \mathbf{I}_3 activate the categories 1, 1, and 2, respectively. Next decrease the vigilance level to 0.8. In response to \mathbf{I}_3 , Category 1 is chosen over Category 2. Indeed

$$\begin{aligned} T_1(\mathbf{I}_3) &= (2|\mathbf{I}_3 \wedge \mathbf{w}_1| - |\mathbf{w}_1| + M(1 - \rho)) \\ &\quad \cdot (|\mathbf{w}_1^c| - M(1 - \rho)) \\ &= (1.78 - 0.9 + 0.2)(1.1 - 0.2) = 0.972 \\ T_2(\mathbf{I}_3) &= (2|\mathbf{I}_3 \wedge \mathbf{w}_2| - |\mathbf{w}_2| + M(1 - \rho)) \\ &\quad \cdot (|\mathbf{w}_2^c| - M(1 - \rho)) \\ &= (2 - 1 + 0.2)(1 - 0.2) = 0.960 \end{aligned}$$

and $T_1(\mathbf{I}_3) > T_2(\mathbf{I}_3)$. Then Category 1 passes the vigilance test

$$\frac{|\mathbf{I}_3 \wedge \mathbf{w}_1|}{M} = 0.89 \geq 0.8$$

and its weight vector becomes $\mathbf{w}_1 = (.39, 0.5)$. So fuzzy ART with the choice function (17) can generalize. ■

APPENDIX E

Lemma 4: Assume fuzzy ART with the choice function (17), and let $0 \leq \rho^- < \rho^+ \leq 1$. By holding trials alternately at ρ^- and at ρ^+ , the number of categories satisfying

$$M(1 - \rho^+) \leq |R_j| \leq M(1 - \rho^-) \quad (39)$$

can increase or remain the same, but never decrease.

Proof: At ρ^+ , a category that satisfies (39) fails the vigilance test, and is therefore conserved. At ρ^- , the same category passes the test, but cannot grow so large as to fail it afterwards. Hence the number of categories satisfying (39) cannot decrease. To show that this number can increase, an example will suffice. Assume that $M = 1, \kappa = 1, \rho^- = 0.9$, and $\rho^+ = 0.95$. First assume $\rho = \rho^-$. In response to $\mathbf{I}_1 = (0.5, 0.5)$, $\mathbf{I}_2 = (0.4, 0.6)$, and $\mathbf{I}_3 = (0.3, 0.7)$, two categories are committed with resulting weight vectors $\mathbf{w}_1 = (0.4, .5)$ and $\mathbf{w}_2 = (0.3, 0.7)$. The inputs $\mathbf{I}_1, \mathbf{I}_2$, and \mathbf{I}_3 activate the categories 1, 1, and 2, respectively. Only Category 1 satisfies

(39). At ρ^+ , two more categories are committed with resulting weight vectors $\mathbf{w}_3 = (0.5, 0.5)$ and $\mathbf{w}_4 = (0.4, 0.6)$. At this point $\mathbf{I}_1, \mathbf{I}_2$, and \mathbf{I}_3 activate the categories 3, 4, and 2, respectively. Back at ρ^- , \mathbf{I}_1 and \mathbf{I}_2 activate Category 1, while \mathbf{I}_3 activates Category 4 with resulting weight vector $\mathbf{w}_4 = (0.3, 0.6)$. As a result, two of the categories, 1 and 4, satisfy (39). ■

APPENDIX F

Using a simple numerical argument, we prove Theorem 1, which states that fuzzy ART with a choice function that meets the fifth requirement of Section III reaches an equilibrium state in finite time for an arbitrary list of inputs, and for an arbitrary list of vigilance levels. Assume an input set containing a finite number K of distinct analog inputs, $\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_K$, that have been complement coded. The proof is based on three lemmas.

Lemma F.1: If $\beta = 1$, then there are 2^K or less possible values for the weight vectors.

Proof: When in fast learning, the weight update equation reduces to (8). Since the fuzzy AND operator is both commutative and associative, the order of the assignments of a category j to the inputs has no effect on the end value of \mathbf{w}_j . In addition, since $\mathbf{I} \wedge \mathbf{I} = \mathbf{I}$, repeated assignments of j to an input leaves \mathbf{w}_j unchanged. The value of \mathbf{w}_j is thus uniquely determined by the assignment or nonassignment of j to the K inputs. From set theory, there are 2^K possible subset choices in a K -element space, and therefore 2^K possible values for \mathbf{w}_j . ■

Lemma F.2: Assume fuzzy ART with the tunable choice function (17). All committed categories have different weight vector values.

Proof: If it were otherwise, then when a category J would be assigned to an input \mathbf{I} , the corresponding weight vector \mathbf{w}_J could take the value of another weight vector \mathbf{w}_j , $J \neq j$. In other words,

$$\mathbf{w}_J^{(\text{new})} = \mathbf{w}_j \quad (40)$$

with $d_j = 0$. Since the choice function is such that if $d_J \neq 0$, then $T_J^{(\text{new})} > T_J^{(\text{old})}$, we have $T_J < T_j$. Therefore J can only be activated if j failed the vigilance test at the current vigilance level. But then

$$\begin{aligned} |R_J| + d_J &= |R_j| \\ &> M(1 - \rho) \end{aligned} \quad (41)$$

and category J also fails the test. Therefore the value of the weight vector of each committed category is unique. ■

Lemmas F.1 and F.2 together guarantee that no more than $(2^K - 1)$ categories can be committed, regardless of the number of times the inputs are presented, and regardless of the vigilance levels. One condition to stability is thus met.

Lemma F.3: Assume Fuzzy ART with the tunable choice function (17). Equilibrium is reached after $2^K(2^K - 1)/2$ or less weight vector changes.

Proof: By virtue of (8), weight vectors cannot cycle through previous values. We have shown in Lemma F.2 that all committed categories have different weight vector values, and in Lemma F.1 that there are no more than 2^K possible

values, including one for uncommitted categories. Therefore w_1 , the weight vector of the first category to be committed, cannot undergo more than $(2^K - 1)$ changes; w_2 more than $(2^K - 2)$ changes, w_3 more than $(2^K - 3)$ changes, and so on. Proceeding inductively, we obtain that the total number of weight vector changes for all possible categories cannot exceed $2^K(2^K - 1)/2$. ■

By definition, every learning trial prior to equilibrium must change at least one weight vector. Since by Lemma F.3 equilibrium is reached after $2^K(2^K - 1)/2$ or less weight vector changes, it is reached after $2^K(2^K - 1)/2$ or less trials, and thus in finite time. Note that this loose upper bound is independent of the vigilance levels used.

ACKNOWLEDGMENT

The authors would like to acknowledge the contribution of E. Granger, who carried out computer simulations for this work.

REFERENCES

- [1] S. Grossberg, "Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors," *Biol. Cybern.*, vol. 23, pp. 121–134, 1976.
- [2] ———, "Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions," *Biol. Cybern.*, vol. 23, pp. 187–202, 1976.
- [3] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, no. 6, pp. 759–771, 1991.
- [4] M. Georgiopoulos, H. Fernlund, G. Bebis, and G. L. Heileman, "Order of search in Fuzzy ART and Fuzzy ARTMAP: Effect of the choice parameter," *Neural Networks*, vol. 9, no. 9, pp. 1541–1559, 1996.
- [5] P. K. Simpson, "Fuzzy min-max neural network—Part 2: Clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 32–44, Feb. 1993.
- [6] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Comput. Vision, Graphics, Image Processing*, no. 37, pp. 54–115, 1987.
- [7] G. Bartfai and R. White, "Adaptive resonance theory-based modular networks for incremental learning of hierarchical clusterings," *Connection Sci.*, vol. 9, no. 1, pp. 87–112, 1997.
- [8] S. Grossberg, Ed., *Neural Networks and Natural Intelligence*. Cambridge, MA: MIT Press, 1988.
- [9] G. A. Carpenter and S. Grossberg, Eds., *Pattern Recognition by Self-Organizing Neural Networks*. Cambridge, MA: MIT Press, 1991.
- [10] ———, "Learning, categorization, rule formation, and prediction by fuzzy neural networks," in *Fuzzy Logic and Neural Network Handbook*, C. Chen, Ed. New York: McGraw-Hill, 1996, ch. 1.
- [11] J. Huang, M. Georgiopoulos, and G. L. Heileman, "Fuzzy ART properties," *Neural Networks*, vol. 8, no. 2, pp. 203–213, 1995.
- [12] B. Moore, "ART 1 and pattern clustering," in *Proc. 1988 Connectionist Models Summer School*, D. Touretzky, G. Hinton, and T. Sejnowski, Eds. San Mateo, CA: Morgan Kaufmann, June 17–26, 1988, pp. 174–185.
- [13] B. Kosko, *Neural Networks and Fuzzy Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [14] G. A. Carpenter and M. N. Gajja, "Fuzzy ART choice functions," Boston Univ. Center Adaptive Syst. Dept. Cognitive Neural Syst., Boston, MA, Tech. Rep. CAS/CNS-93-060, Dec. 1993.
- [15] M. Blume, D. A. Van Blerkom, and S. C. Esener, "Fuzzy ARTMAP modifications for intersecting class distributions," in *Proc. World Conf. Neural Networks*, Sept. 16–18, 1996.
- [16] D. E. Rumelhart and D. Zipser, "Feature discovery by competitive learning," *Cognitive Sci.*, no. 9, pp. 75–112, 1985. Also in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, by D. E. Rumelhart, J. L. McClelland and the PDP Research Group. Cambridge, MA: MIT Press, 1986, ch. 5.
- [17] L. Burke, "Conscientious neural nets for tour construction in the traveling salesman problem: The vigilant net," *Comput. Operations Res.*, vol. 23, no. 2, pp. 121–129, 1995.
- [18] P. Lavoie, J.-F. Crespo, and Y. Savaria, "On the stability of fuzzy ART," in *Proc. 18th Biennial Symp. Commun.*, Queen's Univ., Kingston, Ont., Canada, June 2–5, 1996, pp. 185–188.



Pierre Lavoie (S'85–M'85) received the B.Eng. and Ph.D. degrees in electrical engineering from École Polytechnique de Montréal in 1985 and 1990, respectively.

After these studies he joined the Defence Research Establishment Ottawa, one of the five laboratories of the Canadian Department of National Defense, as a Defense Scientist in the Electronic Support Measures section. His research interests include signal processing algorithms for radar waveform analysis, artificial neural networks for cluster-

ing and classification, and receiver architectures for radar electronic support and electronic countermeasures.

Dr. Lavoie is a National Representative on Panel 4 of the Electronic Warfare Systems group of The Technical Cooperation Program (TTCP), an international organization that collaborates in defense scientific and technical information exchange, program harmonization and alignment, and shared research activities for Australia, Canada, New Zealand, the United Kingdom, and the United States. He has contributed many working papers to TTCP and NATO panels, and published in the open literature eight journal papers and 15 refereed proceedings papers. He is a member of the Association of Old Crows, which supports the electronic defense community.

Jean-François Crespo was born in Montreal, Canada, in 1969. He received the Bachelor's degree in electrical engineering and the Master's degree in electrical engineering from École Polytechnique de Montréal, Canada, in 1992 and 1994, respectively.

Since 1994, he has been a member of the speech research team at Nortel Network's Open Speech business unit. His research interests are speech recognition technology, specifically acoustic modeling, and search techniques.



Yvon Savaria (M'78) received the B.Eng. and M.Sc.A. degrees in electrical engineering from École Polytechnique de Montréal, Canada, in 1980 and 1982, respectively. He also received the Ph.D. degree in electrical engineering in 1985 from McGill University, Montréal.

Since 1985, he has been with École Polytechnique de Montréal. He is currently professor and director of the microelectronics research group. He has carried work in several areas related to microelectronics circuits such as testing,

clocking methods, defect and fault tolerance, high-speed interconnects, and circuit design techniques, CAD methods, reconfigurable computing and applications of microelectronics to telecommunications, image processing video processing, radar signal processing and digital signal processing acceleration. His work in neural networks was driven by the requirements of on-line radar signal processing, where he leads a project on clustering and categorization of radar signals. He has published 36 journal papers and 125 conference papers, and he was the thesis advisor of 52 graduate students that completed their studies.

He was the program Cochairman of the 1992 edition and the Chairman of the 1993 edition of the IEEE Workshop on Defect and Fault Tolerance in VLSI Systems. He has recently been working as a consultant or was sponsored for carrying research by CNRC, Design Workshop, DREO, Genesis, Hyperchip, Miranda, MiroTech, Nortel, and PMC-Sierra. He is a member of the Groupe de Recherche Interuniversitaire en Architecture des Ordinateurs (GRIAO), of the Micronet Canadian center of excellence on microelectronics research and of the Ordre des Ingénieurs du Québec (OIQ).