Fuzzy ART Choice Functions

Gail A. Carpenter and Marin N. Gjaja

Center for Adaptive Systems and Department of Cognitive and Neural Systems
Boston University, 111 Cummington Street, Boston, Massachusetts 02215 USA

Abstract

Adaptive Resonance Theory (ART) models are real-time neural networks for category learning, pattern recognition, and prediction. Unsupervised fuzzy ART and supervised fuzzy ARTMAP networks synthesize fuzzy logic and ART by exploiting the formal similarity between the computations of fuzzy subsethood and the dynamics of ART category choice, search, and learning. Fuzzy ART self-organizes stable recognition categories in response to arbitrary sequences of analog or binary input patterns. It generalizes the binary ART 1 model, replacing the set-theoretic intersection ($\cap$) with the fuzzy intersection ($\wedge$), or component-wise minimum. A normalization procedure called complement coding leads to a symmetric theory in which the fuzzy intersection and the fuzzy union ($\vee$), or component-wise maximum, play complementary roles. A geometric interpretation of fuzzy ART represents each category as a box that increases in size as weights decrease. This paper analyzes fuzzy ART models that employ various choice functions for category selection. One such function minimizes total weight change during learning. Benchmark simulations compare performance of fuzzy ARTMAP systems that use different choice functions.

ART and ARTMAP

Adaptive Resonance Theory (ART) was introduced as a theory of human cognitive information processing (Grossberg, 1976). The theory has led to an evolving series of real-time neural network models for unsupervised and supervised category learning and pattern recognition. These ART models form stable recognition categories in response to arbitrary input sequences with either fast or slow learning. Unsupervised ART networks include ART 1 (Carpenter and Grossberg, 1987a), which stably learns to categorize binary input patterns presented in an arbitrary order; ART 2 (Carpenter and Grossberg, 1987b), which stably learns to categorize either analog or binary input patterns presented in an arbitrary order; and ART 3 (Carpenter and Grossberg, 1990), which carries out parallel search, or hypothesis testing, of distributed recognition codes in a multi-level network hierarchy. Many of the ART papers are collected in the anthology *Pattern Recognition by Self-Organizing Neural Networks* (Carpenter and Grossberg, 1991).

A supervised network architecture, called ARTMAP, self-organizes categorical mappings between m-dimensional input vectors and n-dimensional output vectors. ARTMAP's internal control mechanisms create stable recognition categories of optimal size by maximizing code compression while minimizing predictive error in an on-line setting. Binary ART 1 computations are the foundation of the first ARTMAP network (Carpenter, Grossberg, and Reynolds, 1991), which therefore learns binary maps. Fuzzy ART (Carpenter, Grossberg, and Rosen, 1991) generalizes ART 1 to learn stable recognition categories in response to analog and binary input patterns (Figure 1). The domain of fuzzy ART is thus the same as that of ART 2, but fuzzy ART

Acknowledgments: This research was supported in part by ARPA (ONR N00014-92-J-4015), the National Science Foundation (NSF IRI 90-00530), and the Office of Naval Research (ONR N00014-91-J-4100).

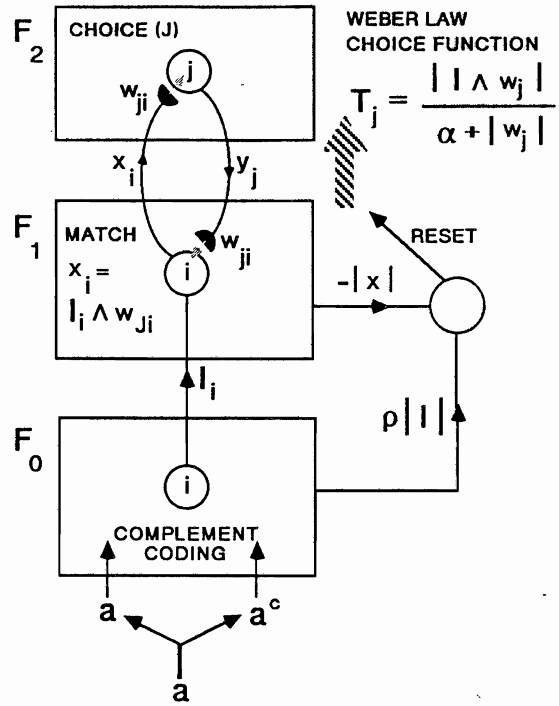

Figure 1: Fuzzy ART module.

measures pattern similarity by the city-block metric, while ART 2 is based on the Euclidean metric. When fuzzy ART replaces ART 1 in an ARTMAP system, the resulting fuzzy ARTMAP architecture (Carpenter, Grossberg, Markuzon, Reynolds, and Rosen, 1992) self-organizes categorical mappings between analog or binary input and output vectors that are stable with fast or slow learning (Figure 2).

This article analyzes fuzzy ART systems that employ various choice functions for category selection. One such network is shown to be optimal in the sense that it minimizes total weight change during learning. Simulations of supervised ARTMAP networks illustrate computational properties of the different fuzzy ART choice functions. The following section outlines the fuzzy ART algorithm with complement coding preprocessing. The limiting case of conservative choice is then examined, along with several alternative choice functions for bottom-up category selection. The function that minimizes the total weight change during learning is more truly conservative than other choice functions. A geometric interpretation of fuzzy ART represents categories as boxes that grow as weights shrink during learning. Benchmark simulations show that alternative choice functions minimally affect system performance. Various choice functions may therefore be selected for their individual computational properties while maintaining the demonstrated utility of ART 1, fuzzy ART, and ARTMAP networks. These studies indicate that when alternative choice functions are selected for reasons such as computational ease or generalizability, the basic ART and ARTMAP dynamics are retained.

Fuzzy ART Algorithm

Normalization by complement coding: Complement coding is a preprocessing step that normalizes fuzzy ART input while preserving amplitude information. When $\mathbf{a} = (a_1, \ldots, a_M)$ is

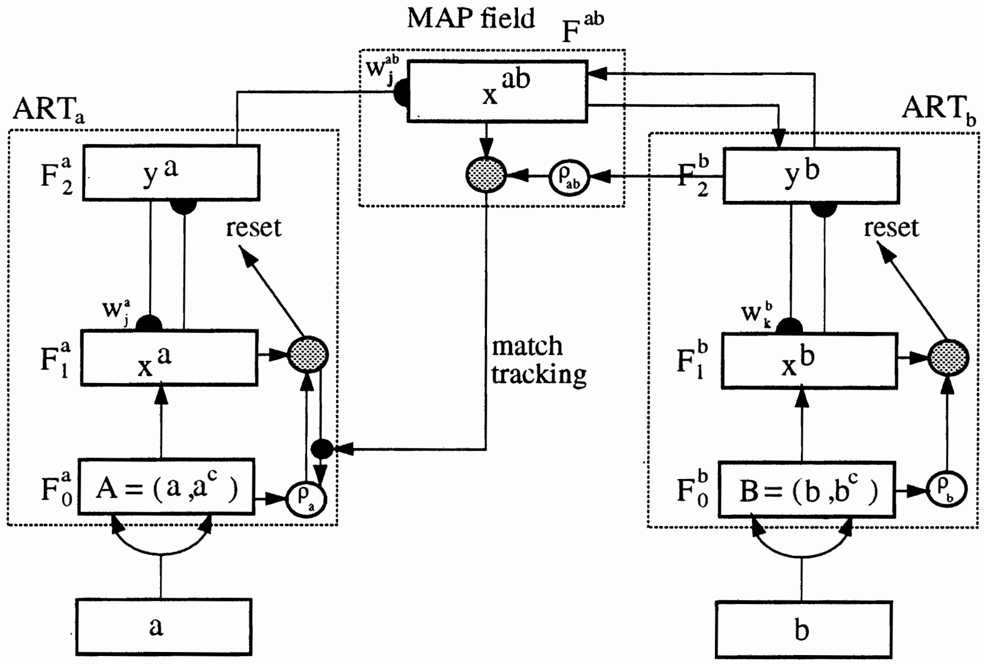

Figure 2: Fuzzy ARTMAP architecture. The ART$_a$ complement coding preprocessor transforms the M_a-vector **a** into the $2M_a$-vector $\mathbf{A} = (\mathbf{a}, \mathbf{a}^c)$ at the ART$_a$ field F_0^a. **A** is the input to the ART$_a$ field F_1^a. Similarly, the input to F_1^b is the $2M_b$-vector $(\mathbf{b}, \mathbf{b}^c)$. When ART$_b$ disconfirms a prediction of ART$_a$, map field inhibition induces the match tracking process. Match tracking raises the ART$_a$ vigilance ρ_a to just above the F_1^a-to-F_0^a match ratio $|\mathbf{x}^a|/|\mathbf{A}|$. This triggers an ART$_a$ search which leads to activation of either an ART$_a$ category that correctly predicts **b** or to a previously uncommitted ART$_a$ category node.

the network input, with $a_i \in [0, 1]$, the complement coded input **I** is the 2M-dimensional vector

$$\mathbf{I} = (\mathbf{a}, \mathbf{a}^c) \equiv (a_1, \ldots, a_M, a_1^c, \ldots, a_M^c), \tag{1}$$

where

$$a_i^c \equiv 1 - a_i \tag{2}$$

(Figure 1). Complement coding implies that $|\mathbf{I}| = M$, with the city-block norm $|\cdot|$ defined by:

$$|\mathbf{I}| \equiv \sum_{i=1}^{2M} I_i. \tag{3}$$

ART field activity vectors: Each ART system includes a field F_0 of nodes that represent a current input vector and a field F_1 that receives both bottom-up input from F_0 and top-down input from a field F_2 that represents the active code, or category. With complement coding, **I** $\equiv (I_1, \ldots, I_{2M})$ denotes F_0 activity, $\mathbf{x} \equiv (x_1, \ldots, x_{2M})$ denotes F_1 activity, and $\mathbf{y} \equiv (y_1, \ldots, y_{2M})$ denotes F_2 activity. The number of nodes in each field can be arbitrarily large.

Weight vector: Associated with each F_2 category node j $(j = 1, \ldots, N)$ is a vector $\mathbf{w}_j \equiv (w_{j1}, \ldots, w_{j,2M})$ of adaptive weights, or long-term memory (LTM) traces. Initially each category is *uncommitted.* After a category codes its first input it becomes *committed.* Each component

w_{ji} can decrease but never increase during learning. Thus each weight vector $\mathbf{w}_j(t)$ converges to a limit. The fuzzy ART weight, or prototype, vector $\mathbf{w}_j$ subsumes both the bottom-up and top-down weight vectors of ART 1.

Initial values: With complement coding, initial values of the weights are:

$$w_{j1}(0) = \ldots = w_{j,2M}(0) = 1. \tag{4}$$

Parameters: A choice parameter $\alpha > 0$, a learning rate parameter $\beta \in [0, 1]$, and a vigilance parameter $\rho \in [0, 1]$ determine fuzzy ART dynamics.

Category choice: The system makes a *category choice* when at most one F_2 node can become active at a given time. A *choice function* $T_j(\mathbf{I})$ determines the selected category. The index J denotes the chosen category, with:

$$T_J = \max\{T_j : j = 1 \ldots N\}. \tag{5}$$

If more than one T_j is maximal, the category with the smallest j index is chosen, so nodes become committed in order $j = 1, 2, 3, \ldots$. When the J^{th} category is chosen, $y_J = 1$ and $y_j = 0$ for $j \neq J$. In a choice system, the F_1 activity vector $\mathbf{x}$ obeys the equation:

$$\mathbf{x} = \begin{cases} \mathbf{I} & \text{if } F_2 \text{ is inactive} \\ \mathbf{I} \wedge \mathbf{w}_J & \text{if the } J^{th}\ F_2 \text{ node is chosen,} \end{cases} \tag{6}$$

where the fuzzy intersection $\wedge$ (Zadeh, 1965) is defined by:

$$(\mathbf{p} \wedge \mathbf{q})_i \equiv \min(p_i, q_i). \tag{7}$$

Weber law choice function: ART 1 (Carpenter and Grossberg, 1987a) and fuzzy ART (Carpenter, Grossberg, and Rosen, 1991) employ a Weber law choice function defined by:

Weber law choice

$$T_j(\mathbf{I}) = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|}, \tag{8}$$

for each F_2 node j.

Resonance or reset: *Resonance* occurs if the chosen category meets the vigilance criterion:

$$|\mathbf{x}| = |\mathbf{I} \wedge \mathbf{w}_J| \geq \rho|\mathbf{I}|. \tag{9}$$

Learning then ensues, as defined below. *Mismatch reset* occurs if:

$$|\mathbf{x}| = |\mathbf{I} \wedge \mathbf{w}_J| < \rho|\mathbf{I}|. \tag{10}$$

Then the value of the choice function T_J is set to 0 for the duration of the input presentation to prevent the persistent selection of the same category during search. A new index J represents the active category, selected again by (5) and (8). The search process continues until the chosen J satisfies the matching criterion (9).

Learning: Once search ends, the weight vector $\mathbf{w}_J$ learns according to the equation:

$$\mathbf{w}_J^{(\text{new})} = \beta(\mathbf{I} \wedge \mathbf{w}_J^{(\text{old})}) + (1 - \beta)\mathbf{w}_J^{(\text{old})}. \tag{11}$$

Fast learning corresponds to setting $\beta = 1$.

Conservative Choice

The linkage between fuzzy subsethood and ART choice/search/learning forms the foundation of the computational properties of fuzzy ART. Vector $\mathbf{w_j}$ is a *fuzzy subset* of $\mathbf{I}$ if:

$$\mathbf{I} \wedge \mathbf{w}_j = \mathbf{w}_j \tag{12}$$

(Zadeh, 1965), i.e., $w_{ji} \leq I_i$ for $i = 1, \ldots, 2M$. When the choice parameter $\alpha = 0^+$, the Weber law choice function $T_j(\mathbf{I})$ (8) measures the *degree* to which $\mathbf{w}_j$ is a fuzzy subset of $\mathbf{I}$ (Kosko, 1986). When $\alpha = 0^+$, $T_j(\mathbf{I})$ is maximized by vectors $\mathbf{w}_j$ that are fuzzy subsets of $\mathbf{I}$, since then $T_j(\mathbf{I}) = 1^-$. A category J for which $\mathbf{w}_J$ is a fuzzy subset of $\mathbf{I}$ will therefore be selected first, if such a category exists. Specifically, the fuzzy subset category J that maximizes $|\mathbf{w}_j|$ will be chosen since then:

$$T_j(\mathbf{I}) = \frac{|\mathbf{w}_j|}{\alpha + |\mathbf{w}_j|}, \tag{13}$$

which is an increasing function of $|\mathbf{w}_j|$. If $\mathbf{w}_J$ is a fuzzy subset of $\mathbf{I}$, learning does not change weights, since then:

$$\mathbf{w}_J^{(\text{new})} = \beta \mathbf{w}_J^{(\text{old})} + (1 - \beta)\mathbf{w}_J^{(\text{old})} = \mathbf{w}_J^{(\text{old})}. \tag{14}$$

Because, when $\alpha = 0^+$, the chosen category J conserves weight values whenever possible, this parameter range is called the fuzzy ART *conservative limit.*

While fuzzy ART choice depends on the degree to which $\mathbf{w}_j$ is a fuzzy subset of $\mathbf{I}$, resonance depends on the degree to which $\mathbf{I}$ is a fuzzy subset of $\mathbf{w}_J$, by (9) and (10). When J is a fuzzy subset choice, then the match function value is:

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} = \frac{|\mathbf{w}_J|}{|\mathbf{I}|}. \tag{15}$$

Choosing J to maximize $|\mathbf{w}_j|$ among fuzzy subset choices thus maximizes the opportunity for resonance in (9). If reset occurs for the node that maximizes $|\mathbf{w}_J|$ among fuzzy subset choices, then reset will also occur for all other subset choices.

Fuzzy ART Choice Functions

The choice function $T_j(\mathbf{I})$ in (8) describes a Weber law form factor that scales the degree of match between the input $\mathbf{I}$ and a weight vector $\mathbf{w}_j$ ($|\mathbf{I} \wedge \mathbf{w}_j|$) relative to the size, or degree of specificity, of $\mathbf{w}_j$. The choice parameter α modulates the scaling process. In the conservative limit, where $\alpha = 0^+$, the rule:

Choice-by-ratio

$$T_j(\mathbf{I}) \cong \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{|\mathbf{w}_j|} \tag{16}$$

determines J, with the largest subset category chosen by (13) when such a category exists. At the opposite extreme, as $\alpha \to \infty$, the rule:

Choice-by-intersection

$$T_j(\mathbf{I}) \cong |\mathbf{I} \wedge \mathbf{w}_j| \tag{17}$$

determines J. Since $\mathbf{I} \wedge \mathbf{w}_j = \mathbf{I}$ for any uncommitted node j, by (4), choice-by-intersection will always select an uncommitted node, unless $\mathbf{I} = \mathbf{w}_J$ for some J. Thus, at this parameter limit, the system's memory consists of exact copies of all input exemplars. As α moves from 0 to ∞, the network becomes progressively more biased in favor of selecting an uncommitted node rather than a coded node with a low match ratio (16). The effect of parametrically raising the choice parameter α from 0 to ∞ is hereby similar to raising the vigilance parameter ρ from 0 to 1.

An alternative to the choice-by-ratio rule (16) *minimizes* the function:

$$T_j(\mathbf{I}) = (|\mathbf{w}_j| - |\mathbf{I} \wedge \mathbf{w}_j|) . \tag{18}$$

This function is related to the membership function used by Simpson (1992). However, Simpson's *fuzzy min-max classifier* does not permit overlapping categories and so does not require a factor, such as (13), to differentiate fuzzy subset categories.

An extension of the rule (18) that is analogous to the Weber law rule (8) minimizes the function $T_j(\mathbf{I})$ defined by:

Choice-by-difference

$$T_j(\mathbf{I}) = (|\mathbf{w}_j| - |\mathbf{I} \wedge \mathbf{w}_j|) + \epsilon\,(|\mathbf{I} \vee \mathbf{w}_j| - |\mathbf{w}_j|) . \tag{19}$$

In (19), $\vee$ denotes the fuzzy union, or component-wise maximum (Zadeh, 1965).

Parameter ϵ in (19) is analogous to the fuzzy ART choice parameter α. When $\epsilon = 0^+$, the category J is chosen to minimize the function (18), unless some $\mathbf{w}_j$ is a fuzzy subset of $\mathbf{I}$. Then, the first term in (19) equals 0, so:

$$T_j(\mathbf{I}) = \epsilon\,(|\mathbf{I} \vee \mathbf{w}_j| - |\mathbf{w}_j|) = \epsilon\,(|\mathbf{I}| - |\mathbf{w}_j|) . \tag{20}$$

The function $T_j(\mathbf{I})$ is therefore minimized by the largest subset category J, if such a category exists. Thus, as in (14), the choice rule approaches a conservative limit as $\epsilon \to 0^+$.

Compared to the Weber law rule (8) with $\alpha = 0^+$, the choice-by-difference rule (19) with $\epsilon = 0^+$ holds a superior claim to the label *conservative.* Both rules make a fuzzy subset choice when possible, so both conserve weights if the fuzzy subset choice J satisfies the vigilance criterion (9). In addition, however, the choice-by-difference rule with $\epsilon = 0^+$ selects the category that *minimizes total weight change* during learning, whether or not $\mathbf{w}_j$ is a fuzzy subset of $\mathbf{I}$, as follows.

Suppose that $\mathbf{w}_j$ is not a fuzzy subset of $\mathbf{I}$. Then choice-by-difference minimizes the function:

$$\begin{aligned} T_j(\mathbf{I}) &= (|\mathbf{w}_j| - |\mathbf{I} \wedge \mathbf{w}_j|) + \epsilon\,(|\mathbf{I} \vee \mathbf{w}_j| - \mathbf{w}_j) \\ &\cong (|\mathbf{w}_j| - |\mathbf{I} \wedge \mathbf{w}_j|) > 0 \end{aligned} \tag{21}$$

when $\epsilon = 0^+$. Suppose that the chosen category J satisfies the vigilance criterion (9). Then the learning law (11) implies that the total weight change during learning is:

$$\begin{aligned} \Delta \mathbf{w}_J &\equiv |\mathbf{w}_J^{(\text{old})} - \mathbf{w}_J^{(\text{new})}| \\ &= \beta\left(|\mathbf{w}_J^{(\text{old})}| - |\mathbf{I} \wedge \mathbf{w}_J^{(\text{old})}|\right) . \end{aligned} \tag{22}$$

Thus selecting J to minimize the choice-by-difference function leads to minimal weight change among all categories that satisfy the vigilance criterion, and to no weight change if $\mathbf{w}_J$ is a fuzzy subset of $\mathbf{I}$.

At the other extreme, as $\epsilon \to \infty$, choice-by-difference minimizes the function:

$$T_j(\mathbf{I}) \cong \epsilon \left(|\mathbf{I} \vee \mathbf{w}_j| - |\mathbf{w}_j| \right). \tag{23}$$

$T_j(\mathbf{I})$ is minimal at uncommitted nodes or when $\mathbf{w}_j = \mathbf{I}$, since then

$$\begin{aligned} T_j(\mathbf{I}) &\cong \epsilon \left(|\mathbf{I} \vee \mathbf{w}_j| - |\mathbf{w}_j| \right) \\ &= \epsilon \left(|\mathbf{w}_j| - |\mathbf{w}_j| \right) = 0. \end{aligned} \tag{24}$$

Thus, like fuzzy ART with Weber law choice as $\alpha \to \infty$, choice-by-difference reduces to exemplar memorization as $\epsilon \to \infty$. Correspondingly, as ϵ moves from 0 to ∞, the degree of code compression generally decreases, as it does when the vigilance parameter ρ moves from 0 to 1.

Alternative choice functions have similar properties in the limit as $\epsilon \to 0^+$. One such function is:

$$T_j(\mathbf{I}) = \left(|\mathbf{w}_j| - |\mathbf{I} \wedge \mathbf{w}_j| \right) + \epsilon \left(|\mathbf{I}| - |\mathbf{w}_j| \right), \tag{25}$$

as in (20). With complement coding, this function is equivalent to:

$$T_j(\mathbf{I}) = \left(|\mathbf{w}_j| - |\mathbf{I} \wedge \mathbf{w}_j| \right) + \epsilon \left(M - |\mathbf{w}_j| \right), \tag{26}$$

since $|\mathbf{I}| \equiv M$. However, choice-by-difference maintains an aesthetic symmetry as well as a form factor that is similar to the difference function that determines resonance (9) or reset (10).

Benchmark simulations will now show that fuzzy ART with the Weber law choice rule (8) has performance characteristics similar to those of a system that is the same except for a choice-by-difference rule (19) determining category selection.

Fuzzy ART Geometry

A geometric interpretation of fuzzy ART represents each category as a box in M-dimensional space, where M is the number of components of input **a**. Consider an input set that consists of 2-dimensional vectors **a**. With complement coding,

$$\mathbf{I} = (\mathbf{a}, \mathbf{a}^c) = (a_1, a_2, 1 - a_1, 1 - a_2). \tag{27}$$

Each category j then has a geometric representation as a rectangle R_j. Following (27), a complement-coded weight vector $\mathbf{w}_j$ takes the form:

$$\mathbf{w}_j = (\mathbf{u}_j, \mathbf{v}_j^c), \tag{28}$$

where $\mathbf{u}_j$ and $\mathbf{v}_j$ are 2-dimensional vectors. Vector $\mathbf{u}_j$ defines the lower left corner of a category rectangle R_j and $\mathbf{v}_j$ defines the upper right corner (Figure 3). The size of R_j is:

$$|R_j| \equiv |\mathbf{v}_j - \mathbf{u}_j|, \tag{29}$$

which is equal to the height plus the width of R_j. In fact, for any M, $|R_j| = M - |\mathbf{w}_j|$.

In a fast-learn fuzzy ART system, with $\beta = 1$ in (11), $\mathbf{w}_J^{(\text{new})} = \mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$ when J is an uncommitted node. The corners of $R_J^{(\text{new})}$ are then $\mathbf{a}$ and $(\mathbf{a}^c)^c = \mathbf{a}$. Hence $R_J^{(\text{new})}$ is just the point **a**. Learning increases the size of R_J, which grows as the size of $\mathbf{w}_J$ shrinks. Vigilance ρ determines the maximum box size, with $|R_j| \leq 2(1 - \rho)$. During each fast-learning trial, R_J

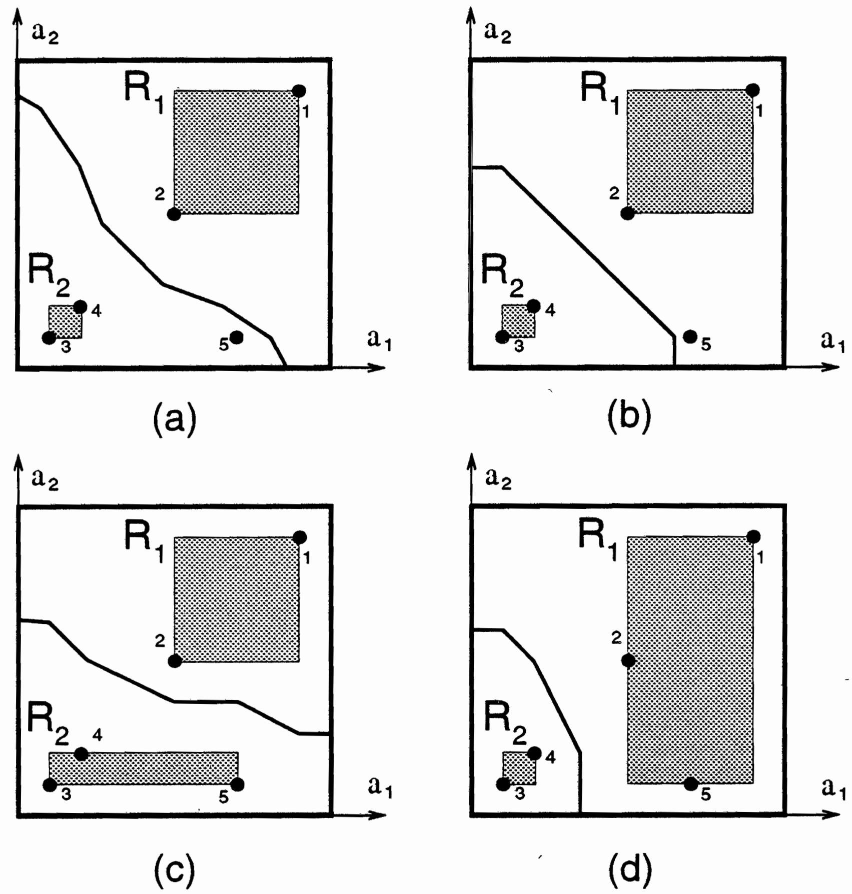

Figure 3: Fuzzy ART category boxes. Simulations (a) and (c) use the Weber law choice function (8), with $\alpha = 0^+$, and (b) and (d) use the choice-by-difference function (19), with $\epsilon = 0^+$. Plots (a) and (b) show category boxes and decision boundaries at the time input 5 is presented. Plots (c) and (d) show the system state after learning. Parameters $\beta = 1.0$ and $\rho = 0.4$.

expands to $R_J^{(\text{old})} \oplus \mathbf{a}$, the minimum rectangle containing $R_J^{(\text{old})}$ and $\mathbf{a}$, with corners $\mathbf{a} \wedge \mathbf{u}_J^{(\text{old})}$ and $\mathbf{a} \vee \mathbf{v}_J^{(\text{old})}$. However, before R_J can expand to include $\mathbf{a}$, reset chooses another category if $|R_J \oplus \mathbf{a}|$ is too large. With fast learning, R_j is the smallest rectangle that encloses all vectors $\mathbf{a}$ that have chosen category j without reset.

Figure 3 illustrates fuzzy ART category boxes at the start (a,b) and end (c,d) of an interval in which input 5 is presented. Plots (a) and (c) use the Weber law choice function (8) and plots (b) and (d) use the choice-by-difference function (19), both in the conservative limit. Vigilance $\rho = 0.4$, so reset occurs if $|R_J \oplus \mathbf{a}| > 1.2$ for a chosen category J. Each plot shows the decision boundary between the set of points $\mathbf{a}$ that would first select box R_1 and the set of points that would select box R_2. In plots (a) and (b), the boxes are the same and the decision boundaries are similar for the two choice functions. However, some points, including input 5, lie on different sides of the boundary. With Weber law choice (a), input 5 chooses $J = 2$, expanding the size of R_2 by 0.5 units during learning (c). With choice-by-difference (b), input 5 chooses $J = 1$, expanding the size of R_1 by 0.4 units during learning (d). This demonstrates the choice-by-difference property of minimal total weight change. Plots (c) and (d) show the different category

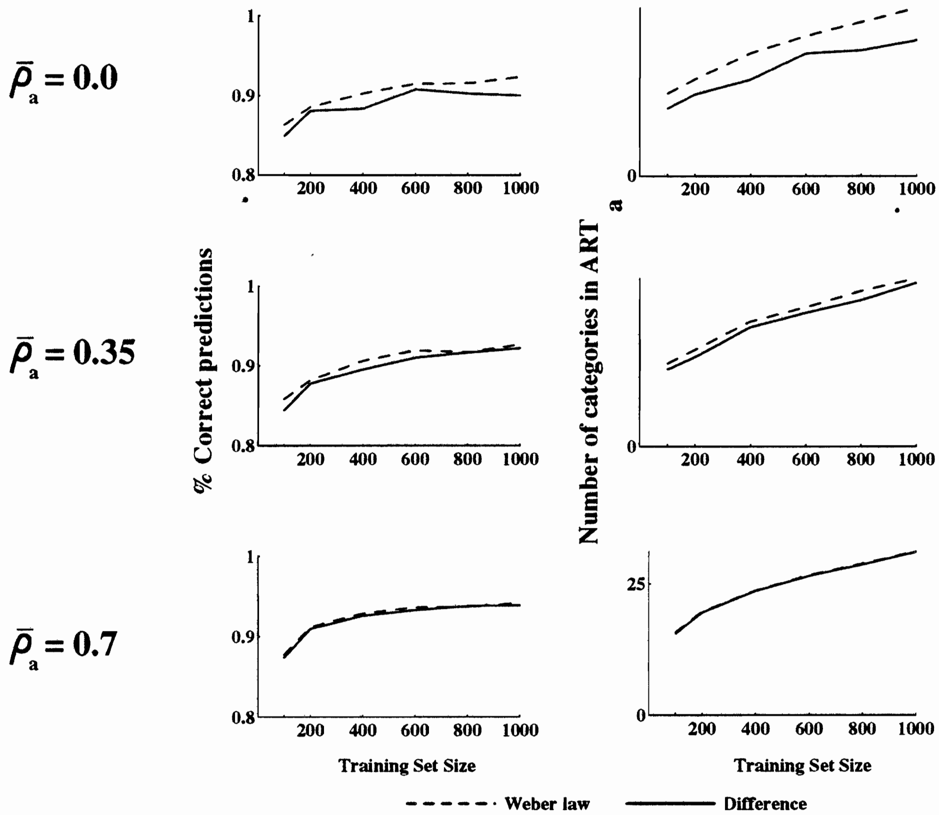

Figure 4: Fuzzy ARTMAP circle-in-the-square simulations for the Weber law choice function (8) (- - - - -) and the choice-by-difference function (19) (——). ART$_a$ baseline vigilance $\bar{\rho}_a$ = 0.0, 0.35, and 0.7, in the conservative limit ($\alpha = 0^+$, $\epsilon = 0^+$), with fast learning ($\beta = 1$).

structures and diverging decision boundaries that can result if a training set input falls near the boundary.

Fuzzy ARTMAP Simulations

An ARPA benchmark simulation, circle-in-the-square (Wilensky, 1990), illustrates fuzzy ARTMAP dynamics. The simulation task is learning to identify which points lie inside and which lie outside a circle. During training, components of the ART$_a$ input **a** are the x- and y-coordinates of a point in the unit square; and ART$_b$ input equals 0 or 1, identifying a as inside or outside the circle. When ARTMAP makes a predictive error during training, *match tracking* raises the ART$_a$ vigilance ρ_a (Figure 2) just enough to trigger search for another F_2^a category. This variable vigilance leads to variable category box sizes as the system balances the competing requirements of code compression (large boxes) and predictive accuracy (small boxes for exceptional cases).

Figure 4 shows fuzzy ARTMAP circle-in-the-square simulation results for the Weber law choice function (dotted lines) and the choice-by-difference function (solid lines), each in the conservative limit with fast learning. Performance is nearly identical for the two choice functions for baseline vigilance parameters $\bar{\rho}_a$ ranging from 0.0 to 0.7 and for training set sizes ranging from 100 to 1000 inputs. Since choice-by-difference minimizes weight change, that system creates slightly fewer categories when $\bar{\rho}_a = 0.0$ and has slightly more test set errors. Even this difference disappears as higher $\bar{\rho}_a$ itself creates more ART_a categories for both choice functions. Similarly, no consistent or significant differences persist for larger values of the choice parameters α and ϵ.

The mushroom database (Schlimmer, 1987) generated the benchmark problem of the original ARTMAP network (Carpenter, Grossberg, and Reynolds, 1991). The Weber law choice function and the choice-by-difference function again show similar performance statistics across a wide range of simulations that use this database. These include on-line and off-line learning with varied baseline vigilance levels and training set sizes.

Performance statistics, plus the added advantage of true conservative learning, argue for the use of the choice-by-difference function (19) when this function has computational properties that are needed for a fuzzy ART network embedded in larger architectures or used for computations beyond the scope of the original system.

References

Carpenter, G.A. and Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, **37**, 54–115.

Carpenter, G.A. and Grossberg, S. (1987b). ART 2: Stable self-organization of pattern recognition codes for analog input patterns. *Applied Optics*, **26**, 4919–4930.

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

Carpenter, G.A. and Grossberg, S. (Eds.) (1991). Pattern Recognition by Self-organizing Neural Networks. Cambridge, MA: MIT Press.

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, **3**, 698–713. Technical Report CAS/CNS-91-016. Boston, MA: Boston University.

Carpenter, G.A., Grossberg, S. and Reynolds, J.H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, **4**, 565–588. Technical Report CAS/CNS-TR-91-001. Boston, MA: Boston University.

Carpenter, G.A., Grossberg, S., and Rosen, D.B. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, **4**, 759–771. Technical Report CAS/CNS-TR-91-015. Boston, MA: Boston University.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, **23**, 187–202.

Kosko, B. (1986). Fuzzy entropy and conditioning. *Information Sciences*, **40**, 165–174.

Schlimmer, J.S. (1987). Mushroom database. UCI Repository of Machine Learning Databases (aha@ics.uci.edu).

Simpson, P. K. (1992). Fuzzy min-max neural networks – Part 1: Classification. *IEEE Transactions on Neural Networks*, **3**, 776-786.

Wilensky, G. (1990). Analysis of neural network issues: Scaling, enhanced nodal processing, comparison with standard classification. DARPA Neural Network Program Review, October 29–30.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, **8**, 338–353.