FUZZY ART: AN ADAPTIVE RESONANCE ALGORITHM FOR RAPID, STABLE CLASSIFICATION OF ANALOG PATTERNS

Gail A. Carpenter, Stephen Grossberg, and David B. Rosen
Center for Adaptive Systems and
Graduate Program in Cognitive and Neural Systems,
Boston University, 111 Cummington Street,
Boston, Massachusetts 02215 USA

Abstract

The Fuzzy ART system introduced herein incorporates computations from fuzzy set theory into ART 1. For example, the intersection ($\cap$) operator used in ART 1 learning is replaced by the MIN operator ($\wedge$) of fuzzy set theory. Fuzzy ART reduces to ART 1 in response to binary input vectors, but can also learn stable categories in response to analog input vectors. In particular, the MIN operator reduces to the intersection operator in the binary case. Learning is stable because all adaptive weights can only decrease in time. A preprocessing step, called complement coding, uses on-cell and off-cell responses to prevent category proliferation. Complement coding normalizes input vectors while preserving the amplitudes of individual feature activations.

1. Summary of the Fuzzy ART Algorithm

Input vector: Each input $\mathbf{I}$ is an M-dimensional vector $(I_1, \ldots, I_M)$, where each component I_i is in the interval $[0, 1]$.

Weight vector: The pattern that defines each category (j) is a weight vector $\mathbf{w}_j \equiv (w_{j1}, \ldots, w_{jM})$. The number (N) of coded categories may be arbitrarily large. Initially

$$w_{j1} = \ldots = w_{jM} = 1, \tag{1}$$

and each category is said to be *uncommitted.* After a node is selected for coding it becomes *committed.* As shown below, each weight component w_{ji} is monotone non-increasing through time and hence converges to a limit. The Fuzzy ART weight vector $\mathbf{w}_j$ subsumes both the bottom-up and top-down weight vectors of ART 1.

Parameters: Fuzzy ART dynamics are determined by a choice parameter $\alpha > 0$; a learning parameter $\beta \in [0, 1]$; and a vigilance parameter $\rho \in [0, 1]$.

Category choice: For each category j, the choice function T_j is defined by

$$T_j = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|}, \tag{2}$$

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

$$(\mathrm{x} \wedge \mathrm{y})_i \equiv \min(x_i, y_i) \tag{3}$$

and where the norm $|\cdot|$ is defined by

$$|\mathrm{x}| \equiv \sum_{i=1}^{M} x_i. \tag{4}$$

0-7803-0164-1/91/0000-0411$01.00©1991 IEEE

The category choice is indexed by J, where

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

If more than one index j gives a maximal T_j, the node with the smallest index is chosen. Thus nodes become committed in order $j = 1, 2, 3, \dots$.

Resonance or reset: *Resonance* occurs if the match function of the chosen node meets the vigilance criterion. That is,

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} \geq \rho. \tag{6}$$

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

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} < \rho. \tag{7}$$

Then the value of the choice function T_J is reset to -1 for the duration of the input presentation. A new index J is chosen, by (5). The search process continues until the chosen J satisfies (6).

Learning: The weight vector $\mathbf{w}_J$ is updated according to the equation

$$\mathbf{w}_J^{(\text{new})} = \beta(\mathbf{I} \wedge \mathbf{w}_J^{(\text{old})}) + (1 - \beta)\mathbf{w}_J^{(\text{old})}. \tag{8}$$

Fast learning corresponds to setting $\beta = 1$. The learning law (8) is the same as that used by Moore (1989).

Fast commitment option: For efficient coding of noisy input sets, it is useful to set $\beta = 1$ when J is an uncommitted node, and then to take $\beta < 1$ after the category is committed. Then $\mathbf{w}_J^{(\text{new})} = \mathbf{I}$ the first time category J becomes active. This option of fast commitment and slow recoding corresponds to ART learning at intermediate rates (Carpenter, Grossberg, and Rosen, 1991; Moore, 1989).

Input normalization option: Moore (1989) describes a category proliferation problem that can occur in some analog ART systems when a large number of inputs erode the norm of weight vectors. Proliferation of categories is avoided in Fuzzy ART if inputs are normalized; that is, if

$$|\mathbf{I}| \equiv \text{constant} \tag{9}$$

for all inputs $\mathbf{I}$. Normalization can be achieved by preprocessing each incoming vector $\mathbf{a}$, for example setting

$$\mathbf{I} = \frac{\mathbf{a}}{|\mathbf{a}|}. \tag{10}$$

An alternative device, called *complement coding*, achieves normalization while preserving amplitude information. Complement coding represents both the on-response and the off-response to $\mathbf{a}$. To define such a code in its simplest form, let $\mathbf{a}$ itself represent the on-response. The complement of $\mathbf{a}$, denoted by $\mathbf{a}^c$, represents the off-response, where

$$a_i^c \equiv 1 - a_i. \tag{11}$$

The input $\mathbf{I}$ to the recognition system is the 2M-dimensional vector

$$\mathbf{I} \equiv (\mathbf{a}, \mathbf{a}^c) \equiv (a_1, \ldots, a_M, a_1^c, \ldots, a_M^c). \tag{12}$$

Note that

$$|\mathbf{I}| = |(\mathbf{a}, \mathbf{a}^c)| = M, \tag{13}$$

so inputs preprocessed into complement coding form are automatically normalized. Initially, $w_{j1} = \ldots = w_{j,2M} = 1$.

2. Fuzzy Subset Choice

In fast-learn ART 1, if the choice parameter α in (2) is close to 0, the first category chosen by (5) will always be the category whose weight vector $\mathbf{w}_J$ is the largest coded subset of the input vector $\mathbf{I}$, if such a category exists (Carpenter and Grossberg, 1987). Moreover, when $\mathbf{w}_J$ is a subset of $\mathbf{I}$ during resonance, $\mathbf{w}_J$ is unchanged, or conserved, during learning. Thus the limit $\alpha \to 0$ may be called the *conservative limit*, since small values of α tend to minimize recoding during learning. For analog vectors, the degree to which $\mathbf{x}$ is a fuzzy subset of $\mathbf{y}$ is given by the term

$$\frac{|\mathbf{x} \wedge \mathbf{y}|}{|\mathbf{x}|} \tag{14}$$

(Kosko, 1986; Zadeh, 1965). Thus, in the conservative limit, the choice function T_j in (2) reflects the degree to which the weight vector $\mathbf{w}_j$ is a fuzzy subset of the input vector $\mathbf{I}$. If

$$\frac{|\mathbf{I} \wedge \mathbf{w}_j|}{|\mathbf{w}_j|} = 1 \tag{15}$$

category j is said to be a *fuzzy subset choice* for input $\mathbf{I}$. In this case, by (8), no recoding occurs if j is selected since $(\mathbf{I} \wedge \mathbf{w}_j) = \mathbf{w}_j$.

Conversely, in (6), resonance depends on the degree to which $\mathbf{I}$ is a fuzzy subset of $\mathbf{w}_J$. Moreover, if category j is a subset choice, the match function value is given by

$$\frac{|\mathbf{I} \wedge \mathbf{w}_j|}{|\mathbf{I}|} = \frac{|\mathbf{w}_j|}{|\mathbf{I}|}. \tag{16}$$

Thus, choosing J to maximize $|\mathbf{w}_j|$ among fuzzy subset choices also maximizes the opportunity for resonance. If reset occurs for the node that maximizes $|\mathbf{w}_j|$, reset will also occur for all other subset choices.

3. A Geometric Interpretation of Fuzzy ART

Consider a fast-learn Fuzzy ART system; i.e., set $\beta = 1$ in (8). Let the input set consist of 2-dimensional vectors $\mathbf{a}$ in complement coding form. Thus

$$\mathbf{I} = (\mathbf{a}, \mathbf{a}^c) = (a_1, a_2, 1 - a_1, 1 - a_2). \tag{17}$$

In this case, each category j has a geometric representation as a rectangle R_j. If a new vector $\mathbf{a}$ activates a category j and if $\mathbf{a}$ lies within R_j, neither reset nor weight change occurs. If $\mathbf{a}$ lies outside R_j, R_j will expand to include $\mathbf{a}$, unless this expansion would cause R_j to

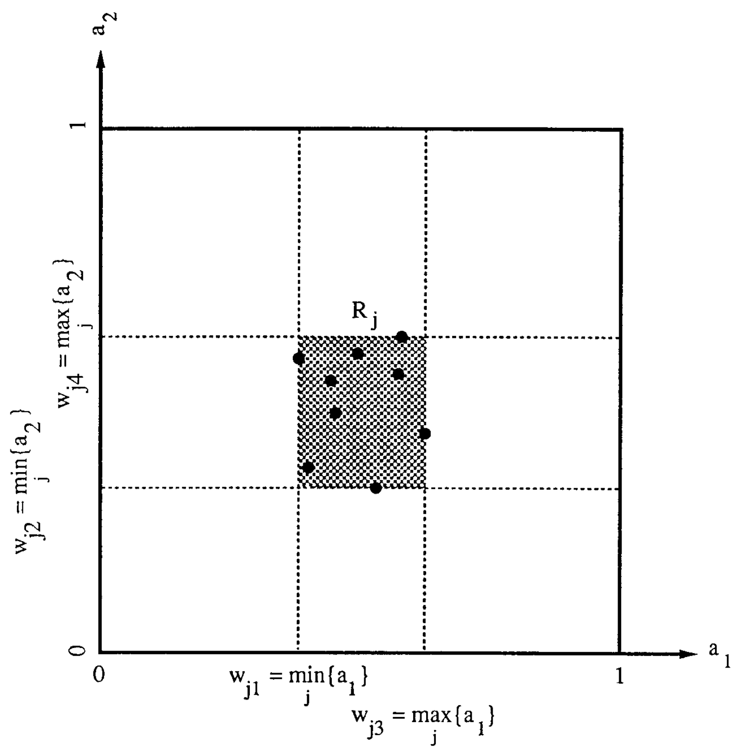

Figure 1. R_j includes all those vectors **a** in the unit square which, in complement coding form, have activated category j. The matching criterion (6) is equivalent to the constraint that the sum of the width and length of R_j be less than $2(1-\rho)$.

become too large. In fact, the length of the horizontal plus the vertical sides of the expanded rectangle must remain less than $2(1-\rho)$. If an expanded R_j would fail to meet this criterion, another category becomes active. Large R_j correspond to small $|\mathbf{w}_j|$. The matching criterion (6) sets an upper bound on the size of R_j.

In general, if **a** has dimension M, the hypercube R_j includes the two vertices $\wedge_j \mathbf{a}$ and $\vee_j \mathbf{a}$, where

$$(\wedge_j \mathbf{a})_i = \min\{a_i : \mathbf{a} \text{ has been coded by category } j\} \equiv \min_j\{a_i\} \tag{18}$$

and

$$(\vee_j \mathbf{a})_i = \max\{a_i : \mathbf{a} \text{ has been coded by category } j\} \equiv \max_j\{a_i\} \tag{19}$$

(Figure 1). The size of R_j is given by

$$|R_j| = \sum_{i=1}^{M}[\max_j\{a_i\} - \min_j\{a_i\}] = [|\vee_j \mathbf{a}| - |\wedge_j \mathbf{a}|]. \tag{20}$$

By (8),

$$\mathbf{w}_j = (\min_j\{a_1\}, \ldots, \min_j\{a_M\}, 1 - \max_j\{a_1\}, \ldots, 1 - \max_j\{a_M\}), \tag{21}$$

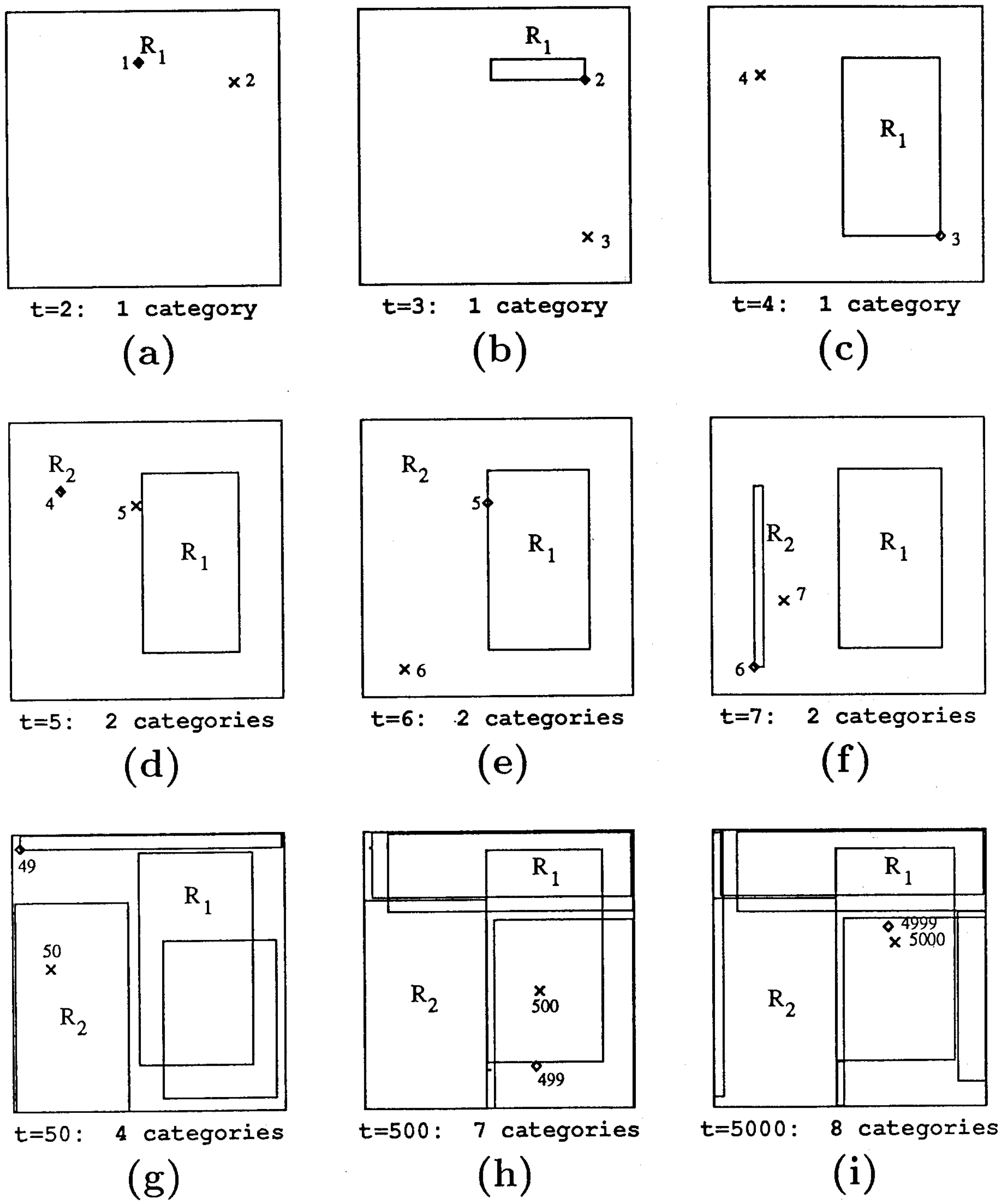

Figure 2. Fuzzy ART simulation with $\alpha \cong 0$, $\beta = 1$, $\rho = .4$, and input vectors a distributed randomly in the unit square. Rectangles R_j grow during learning, and new categories are established, until the entire square is covered.

so

$$|\mathbf{w}_j| = M - \sum_{i=1}^{M}[\max_j\{a_i\} - \min_j\{a_i\}] = M - [|\vee_j \mathbf{a}| - |\wedge_j \mathbf{a}|]. \tag{22}$$

The size of the hypercube R_j is thus

$$|R_j| = [M - |\mathbf{w}_j|]. \tag{23}$$

By (6) and (8),

$$|\mathbf{w}_j| \geq \rho M, \tag{24}$$

since $|\mathbf{I}| \equiv M$, so

$$|R_j| \leq M(1-\rho). \tag{25}$$

Thus high vigilance ($\rho \cong 1$) leads to small R_j while low vigilance ($\rho \cong 0$) permits large R_j. If j is an uncommitted node, $|\mathbf{w}_j| = 2M$ and so, by (23), $|R_j| \equiv -M$.

In the simulation illustrated in Figure 2, vectors $\mathbf{a}$ ($M = 2$) are randomly distributed in the unit square. Each frame shows the new vector $\mathbf{a}^{(t)}$ (cross) and the previous vector $\mathbf{a}^{(t-1)}$ (diamond), as well as the set of rectangles R_j present before learning occurs. The system is run in the fast-learn, conservative limit, and $\rho = .4$. When a new category is established, R_j is just a point, since $\wedge_j \mathbf{a} = \vee_j \mathbf{a} = \mathbf{a}$. If $\mathbf{a}$ lies within one or more established R_j, the rectangle chosen is the one that has the smallest size $|R_j|$. In this case, neither reset nor weight change occurs. If $\mathbf{a}$ does not lie within any established R_j, each new input that activates category j expands R_j unless (as in (d)) such an expansion would cause the size of R_j to exceed $2(1-\rho) = 1.2$. Asymptotically, all points of the square are covered by the set of rectangles. Thereafter no weight changes occur.

Acknowledgements: Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0175 and AFOSR 90-0128), the Army Research Office (ARO DAAL-03-88-K0088), British Petroleum (89-A-1204), DARPA (AFOSR 90-0083), and the National Science Foundation (NSF IRI-90-00530). The authors wish to thank Cynthia E. Bradford for her valuable assistance in the preparation of the manuscript.

References

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

Carpenter, G.A., Grossberg, S., and Rosen, D.B. (1991). ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, **4**, in press.

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

Moore, B. (1989). ART 1 and pattern clustering. In D. Touretzky, G. Hinton, and T. Sejnowski (Eds.), **Proceedings of the 1988 Connectionist Models Summer School**. San Mateo, CA: Morgan Kaufmann Publishers.

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