Pergamon

*Neural Networks*, Vol. 9, No. 9, pp. 1541–1559, 1996

Printed in Great Britain
0893–6080/96 $15.00+.00

PII: S0893-6080(96)00018-4

*CONTRIBUTED ARTICLE*

# Order of Search in Fuzzy ART and Fuzzy ARTMAP: Effect of the Choice Parameter

MICHAEL GEORGIOPOULOS,[1] HANS FERNLUND,[2] GEORGE BEBIS[1] AND GREGORY L. HEILEMAN[3]

[1]University of Central Florida, [2]Dalarna University College, and [3]University of New Mexico

(*Received* 7 *September* 1995; *revised and accepted* 24 *January* 1996)

**Abstract**—*This paper focuses on two ART architectures, the Fuzzy ART and the Fuzzy ARTMAP. Fuzzy ART is a pattern clustering machine, while Fuzzy ARTMAP is a pattern classification machine. Our study concentrates on the order according to which categories in Fuzzy ART, or the* $ART_a$ *model of Fuzzy ARTMAP are chosen. Our work provides a geometrical, and clearer understanding of why, and in what order, these categories are chosen for various ranges of the choice parameter of the Fuzzy ART module. This understanding serves as a powerful tool in developing properties of learning pertaining to these neural network architectures; to strengthen this argument, it is worth mentioning that the order according to which categories are chosen in ART 1 and ARTMAP provided a valuable tool in proving important properties about these architectures. Copyright © 1996 Elsevier Science Ltd.*

**Keywords**—Neural network, Clustering, Classification, Learning, Adaptive resonance theory, Fuzzy ART, Fuzzy ARTMAP.

## 1. INTRODUCTION

In this paper we focus our attention on two ART architectures, *Fuzzy ART* and *Fuzzy ARTMAP*. These architectures were introduced by Carpenter et al. (1991b, 1992), and they belong to the class of neural network architectures that fall under the category of *adaptive resonance theory* (ART) neural networks. Adaptive resonance theory was developed by Grossberg (1976), and a list of some of the ART architectures introduced in the last 10 years are included in the reference list (Carpenter & Grossberg, 1987a, b, 1990; Carpenter et al., 1991a, b, 1992; Healy et al., 1993).

Fuzzy ART is a pattern clustering machine that is capable of clustering arbitrary collections of arbitrarily complex analog input patterns, while Fuzzy ARTMAP is a pattern classification machine that is capable of establishing an arbitrary mapping between an arbitrary collection of analog input patterns and an arbitrary collection of corresponding analog output patterns. Fuzzy ARTMAP has been successfully used to solve pattern classification problems in a fast and efficient way (Carpenter et al., 1992).

It is worth mentioning that Fuzzy ART is one of the components of a Fuzzy ARTMAP architecture, in the sense that Fuzzy ARTMAP consists of two Fuzzy ART modules and an inter-ART module. One of the Fuzzy ART modules receives as inputs the input patterns of the pattern classification task, and the other Fuzzy ART module receives as inputs the corresponding output patterns of the pattern classification task. The Fuzzy ART module that accepts the input patterns clusters them into appropriate categories, while the Fuzzy ART module that accepts the output patterns also clusters them into appropriate categories. Our main focus in this paper is the operation of Fuzzy ARTMAP for pattern classification problems, which are many-to-one maps. Hence, the kind of clustering done at the Fuzzy ART module that accepts as inputs the output patterns is trivial, because every output pattern forms its own cluster. The type of clustering established in the Fuzzy ART module that accepts as inputs the input patterns is much more interesting. From this perspective Fuzzy ART and Fuzzy ARTMAP can be jointly investigated.

In this paper, we focus our attention on the clusters established by Fuzzy ART, or the Fuzzy ART module of Fuzzy ARTMAP that accepts as

Acknowledgements: This research was supported in part by a grant from Boeing Computer Services under Contract W-300445.

Requests for reprints should be sent to Michael Georgiopoulos, Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL 32816, USA; Tel.: (407) 823-5338.

inputs the input patterns of the pattern classification task. In particular, we investigate the order according to which these clusters are chosen when an input pattern is presented to the Fuzzy ART module. In this investigation, we will emphasize the geometrical interpretation of the clusters (categories) formed in the Fuzzy ART module, because this interpretation gives a better understanding of how categories are chosen by Fuzzy ART. To the best of our knowledge, such an investigation has been fully conducted for ART1 (Carpenter & Grossberg, 1987b), and partially for Fuzzy ART (Carpenter et al., 1991b; Huang et al., 1995) and ARTMAP (Georgiopoulos et al., 1994).

The organization of the paper is as follows: in Section 2, we discuss the specifics of the Fuzzy ART architecture. The reader who is familiar with Fuzzy ART can go over this section very quickly, paying attention only to the geometrical interpretation of templates in Fuzzy ART and the distance of a pattern from a template. In Section 3, we describe the major results pertaining to the order of search in Fuzzy ART. Specifically, we first present three fundamental theorems, and then we state three results that describe how Fuzzy ART is making category choices for three distinct ranges of the choice parameter values. The choice parameter is a parameter of the Fuzzy ART architecture. In Section 4, we elaborate on our results by providing some order of search rules, with accompanying figures, that illustrate how Fuzzy ART chooses categories to represent the input patterns presented to it. In Section 5 we extend the Fuzzy ART results of Sections 3 and 4 to the Fuzzy ARTMAP architecture. In Section 6, we make some important observations pertaining to the results developed, and finally, in Section 7, we provide a short review.

## 2. FUZZY ART

### 2.1. Fuzzy ART Architecture

The Fuzzy ART neural network architecture is shown in Figure 1. It consists of two subsystems, the *attentional subsystem*, and the *orienting subsystem*. The attentional subsystem consists of two fields of nodes denoted $F_1^a$ and $F_2^a$. The $F_1^a$ field is called the *input field* because input patterns are applied to it. The $F_2^a$ field is called the *category* or *class representation field* because it is the field where category representations are formed. These category representations represent the clusters to which the input patterns, presented at the $F_1^a$ field, belong. The orienting subsystem consists of a single node (called the *reset node*), which accepts inputs from the $F_1^a$ field, the $F_2^a$ field (not shown in Figure 1), and the input pattern applied across the $F_1^a$ field. The output of the reset node affects the nodes of the $F_2^a$ field.

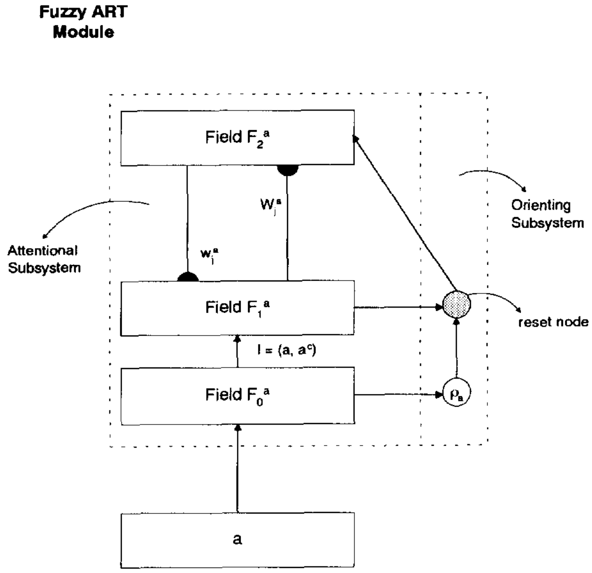


FIGURE 1. A block diagram of the Fuzzy ART architecture.

Some preprocessing of the input patterns of the pattern clustering task takes place before they are presented to Fuzzy ART. The first preprocessing stage takes as an input an $M_a$-dimensional input pattern from the pattern clustering task and transforms it into an output vector $\mathbf{a} = (a_1, \ldots, a_{M_a})$, whose every component lies in the interval [0, 1] (i.e., $0 \leqslant a_i \leqslant 1$ for $1 \leqslant i \leqslant M_a$). The second preprocessing stage accepts as an input the output $\mathbf{a}$ of the first preprocessing stage and produces an output vector $\mathbf{I}$, such that

$$\mathbf{I} = (\mathbf{a}, \mathbf{a}^c) = \left(a_1, \ldots, a_{M_a}, a_1^c, \ldots, a_{M_a}^c\right) \tag{1}$$

where

$$a_i^c = 1 - a_i;\ 1 \leqslant i \leqslant M_a. \tag{2}$$

The above transformation is called *complement coding*. The complement coding operation is performed in Fuzzy ART at a preprocessor field designated by $F_0^a$ (see Figure 1). From now on, we will be referring to the vector $\mathbf{I}$ as the *input pattern*.

We denote a node in the $F_1^a$ field by the index $i (i \in \{1, 2, \ldots, 2M_a\})$, and a node in the $F_2^a$ field by the index $j (j \in \{1, 2, \ldots, N_a\})$. Every node $i$ in the field $F_1^a$ field is connected via a bottom-up weight with every node $j$ in the $F_2^a$ field; this weight is denoted by $W_{ij}^a$. Also, every node $j$ in the $F_2^a$ field is connected via a top-down weight with every node $i$ in the $F_1^a$ field; this weight is denoted by $w_{ji}^a$. The vector whose components are equal to the top-down weights emanating from node $j$ in the $F_2^a$ field is designated by $\mathbf{w}_j^a$ and it is called a *template*. Note that $\mathbf{w}_j^a = (w_{j1}^a, w_{j2}^a, \ldots, w_{j,2M_a}^a)$ for $j = 1, \ldots, N_a$. The vector of bottom-up weights converging to a node $j$ in the $F_2^a$ field is designated by $\mathbf{W}_j^a$. Note that $\mathbf{W}_j^a = (W_{1,j}^a, W_{2,j}^a, \ldots, W_{2M_a,j}^a)$ for $j = 1, \ldots, N_a$. Initial values of the bottom-up and top-down weights are designated by $W_{ij}^a(0)$, and $w_{ji}^a(0)$, respectively. Initial values of the top-down weights are chosen equal to one. Initial values of the bottom-up weights are chosen equal to:

$$\frac{1}{\alpha_a + M_u^a} \tag{3}$$

where $\alpha_a$ and $M_u^a$ are Fuzzy ART parameters. The parameter $\alpha_a$ is called the *choice parameter* and it takes values in the interval $(0, \infty)$. The parameter $M_u^a$ takes values in the interval $[2M_a, \infty)$; we name this parameter *uncommitted node choice parameter*. The initial bottom-up and top-down weight choices in Fuzzy ART correspond to the values of these weights prior to presentation of any input pattern to the Fuzzy ART architecture.

In the original Fuzzy ART paper (Carpenter et al., 1991b) only the top-down weights of the architecture are introduced. We have followed a different approach in this paper, introducing both bottom-up and top-down weights, so that we can naturally introduce the *uncommitted node choice parameter* $M_u^a$ which plays a significant role on the order according to which nodes are chosen in the $F_2^a$ field of Fuzzy ART. Note also, that in the ART 1 paper (Carpenter & Grossberg, 1987b) both bottom-up and top-down weights were introduced.

Before we proceed with the rest of our work it is important to clarify the notation $\mathbf{w}_j^{a,old}$ and $\mathbf{W}_j^{a,old}$, or the notation $\mathbf{w}_j^{a,new}$ and $\mathbf{W}_j^{a,new}$, which is used extensively throughout the paper. Quite often, templates and bottom-up weights in Fuzzy ART are discussed with respect to an input pattern $\mathbf{I}$ presented at the $F_1^a$ field of Fuzzy ART. In particular, the notation $\mathbf{w}_j^{a,old}$ or $\mathbf{W}_j^{a,old}$ denotes the template of node $j$ or the bottom-up weight converging to node $j$ in the $F_2^a$ field of Fuzzy ART, *prior* to the presentation of an input pattern $\mathbf{I}$ at the $F_1^a$ field. Furthermore, the notation $\mathbf{w}_j^{a,new}$ or $\mathbf{W}_j^{a,new}$ denotes the template of node $j$ or the bottom-up weight converging to node $j$ in the $F_2^a$ field of Fuzzy ART, *after* the presentation of an input pattern $\mathbf{I}$ at the $F_1^a$ field. Similarly, any other quantities defined with a superscript {a, old} or {a, new} will indicate values of these quantities prior to and after a pattern presentation to Fuzzy ART, respectively.

## 2.2. Operation of Fuzzy ART

We have used $\mathbf{I}$ to indicate an input pattern applied at the $F_1^a$ field and $\mathbf{w}_j^a$ to indicate the template of node $j$ in $F_2^a$. We will use $|\mathbf{I}|$ and $|\mathbf{w}_j^a|$ to denote the size of $I$ and $\mathbf{w}_j^a$, respectively. The size of a vector in Fuzzy ART is defined to be the sum of its components. Furthermore we define $\mathbf{I} \wedge \mathbf{w}_j^a$ to be the vector whose $i$th component is the minimum of the $i$th $\mathbf{I}$ component and the $i$th $\mathbf{w}_j^a$ component. The operation $\wedge$ is called the *fuzzy-min* operation, while a related operation designated by $\vee$ is called the *fuzzy-max* operation. These operations are shown in Figure 2 for two two-dimensional vectors, denoted by $\mathbf{x}$ and $\mathbf{y}$.

Let us assume that an input pattern $\mathbf{I}$ is presented at the $F_1^a$ field of Fuzzy ART. The appearance of pattern $\mathbf{I}$ across the $F_1^a$ field produces bottom-up inputs that affect the nodes in the $F_2^a$ field. These bottom-up inputs are given by the equation:

$$T_j^a(\mathbf{I}) = \begin{cases} \dfrac{|\mathbf{I}|}{\alpha_a + M_u^a} & \text{if } j \text{ is an uncommitted node} \\ |\mathbf{I} \wedge \mathbf{w}_j^{a,old}| / (\alpha_a + |\mathbf{w}_j^{a,old}|) & \text{if } j \text{ is a committed node} \end{cases} \tag{4}$$

where $\alpha_a$ and $M_u^a$ are Fuzzy ART parameters mentioned previously. A node in $F_2^a$ is called an *uncommitted node* if all of its top-down weights are equal to the initial top-down weights values (i.e., equal to one); otherwise the node is called a *committed node*.

The bottom-up inputs activate a competition process among the $F_2^a$ nodes, which eventually leads to the activation of a single node in $F_2^a$, namely the node which receives the maximum bottom-up input

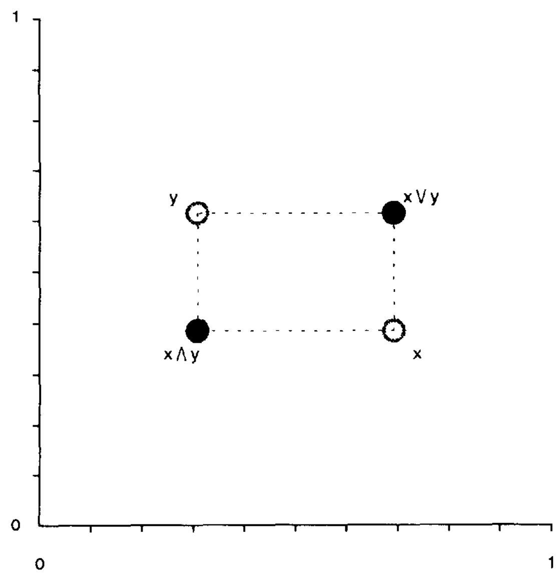


**FIGURE 2. Illustration of the fuzzy min (∧) and the fuzzy max (∨) operations in two-dimensional space. The horizontal axis and the vertical axes designate the first and the second component of the two-dimensional vectors, respectively.**

from $F_1^a$. Let us assume that node $J$ in $F_2^a$ has been activated through this process. The activation of node $J$ in $F_2^a$ indicates that this node is considered as a potential candidate by Fuzzy ART to represent the input pattern **I**. The appropriateness of this node is checked by examining the ratio

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J^{a,old}|}{|\mathbf{I}|}. \tag{5}$$

If this ratio is smaller than $\rho_a$, then node $J$ is deemed inappropriate to represent the input pattern **I**, and as a result it is reset (deactivated). The parameter $\rho_a$ is called the *vigilance parameter* and it takes values in the interval [0, 1]. The deactivation process is carried out by the orienting subsystem, and in particular by the reset node. If a reset happens, another node in $F_2^a$ (different from node $J$) is chosen to represent the input pattern **I**; this resets last for the entire input pattern presentation. The above process continues until an appropriate node in $F_2^a$ is found, or until all the nodes in $F_2^a$ are exhausted. If a node in $F_2^a$ is found appropriate to represent the input pattern **I**, then learning ensues according to the following rules.

Assuming that node $J$ has been chosen to represent **I**, the corresponding top-down weight vector $\mathbf{w}_J^{a,old}$ becomes equal to $\mathbf{w}_J^{a,new}$, where

$$\mathbf{w}_J^{a,new} = \left(\mathbf{I} \wedge \mathbf{w}_J^{a,old}\right). \tag{6}$$

Also, the corresponding bottom-up weight vector $\mathbf{W}_J^{a,old}$ becomes equal to $\mathbf{W}_J^{a,new}$, where

$$\mathbf{W}_J^{a,new} = \frac{\mathbf{I} \wedge \mathbf{w}_J^{a,old}}{\alpha_a + |\mathbf{I} \wedge \mathbf{w}_J^{a,old}|}. \tag{7}$$

It is worth mentioning that in eqns (6) and (7) we might have $\mathbf{w}_J^{a,new} = \mathbf{w}_J^{a,old}$ and $\mathbf{W}_J^{a,new} = \mathbf{W}_J^{a,old}$; in this case we say that no learning occurs for the weights of node $J$. Also note that eqns (6) and (7) are a special case of the learning equations of Fuzzy ART [see Carpenter et al. (1991b)]. This special case of learning is referred to as *fast learning*. The results in this paper are valid only for the fast learning case.

We say that node $J$ has *coded* input pattern **I** if during **I**'s presentation at $F_1^a$, node $J$ in $F_2^a$ is chosen to represent **I**, and **J**'s bottom-up and top-down weights are modified, as eqns (6) and (7) prescribe.

Note that the weights converging to or emanating from an $F_2^a$ node other than $J$ (i.e., the chosen node) remain unchanged during **I**'s presentation.

## 2.3. Operating Phases of Fuzzy ART

Fuzzy ART may operate in two different phases: the *training phase*, and the *performance phase*. Furthermore, in the training phase Fuzzy ART may operate under two different scenarios.

The *first scenario* of the training phase is as follows: we have a collection of input patterns designated as $\mathbf{I}^1, \mathbf{I}^2, \ldots, \mathbf{I}^P$ that we refer to as the *training list*. We want Fuzzy ART to cluster these input patterns into different categories. Obviously, we expect patterns that are similar to each other to be clustered in the same category by Fuzzy ART. In order to achieve the aforementioned goal, we present the training list repeatedly to the Fuzzy ART architecture. That is, we present $\mathbf{I}^1$, then $\mathbf{I}^2$, and eventually $\mathbf{I}^P$; this corresponds to one *list presentation*. We present the training list as many times as it is necessary for Fuzzy ART to cluster the input patterns. The clustering task is considered accomplished (i.e., the learning is complete) if the weights in the Fuzzy ART architecture do not change during a list presentation. The aforementioned training scenario is called *off-line training*.

In the *second scenario* of training we do not have a predetermined list of input patterns that are to be clustered. On the contrary, input patterns are presented one-by-one to the Fuzzy ART architecture, and in most instances patterns are not presented more often than once. Fuzzy ART learns these patterns according to the Fuzzy ART rules described in the previous sections, and training is over when the learning process is disengaged. This second scenario of training is called *on-line training*.

In the *performance phase* of Fuzzy ART the learning process is disengaged and patterns from a *test list* are presented in order to evaluate the clustering performance of Fuzzy ART. Specifically, an input pattern from the test list is presented to Fuzzy ART and through the Fuzzy ART rules, discussed previously, a node $J$ is chosen in $F_2^a$ that is found appropriate to represent the input pattern from the test list. Assuming that some criteria exist for determining how well node $J$ represents the cluster to which the input pattern presented to Fuzzy ART belongs, we can apply this process to all the input patterns from the test list to determine how well Fuzzy ART clusters them. Of course, our results are heavily dependent on the criteria used to judge the clustering performance of Fuzzy ART.

## 2.4. Templates in Fuzzy ART: A Geometrical Interpretation

We previously referred to the top-down weights emanating from a node in the $F_2^a$ field as a *template*. A template corresponding to a committed node is called a *committed template*, while a template corresponding to an uncommitted node is called an *uncommitted template*. As we have already mentioned, an uncommitted template has all of its components equal to one.

In the original Fuzzy ART paper (Carpenter et al., 1991b) it is demonstrated that a committed template $\mathbf{w}_j^a$, which has coded input patterns $\mathbf{I}^1 = (\mathbf{a}(1), \mathbf{a}^c(1))$, $\mathbf{I}^2 = (\mathbf{a}(2), \mathbf{a}^c(2)), \ldots, \mathbf{I}^P = (\mathbf{a}(P), \mathbf{a}^c(P))$, can be written as follows:

$$\mathbf{w}_j^a = \mathbf{I}^1 \wedge \mathbf{I}^2 \wedge \ldots \wedge \mathbf{I}^P = \left(\wedge_{i=1}^P \mathbf{a}(i), \wedge_{i=1}^P \mathbf{a}^c(i)\right) = \left(\wedge_{i=1}^P \mathbf{a}(i), \{\vee_{i=1}^P \mathbf{a}(i)\}^c\right) \quad (8)$$

or

$$\mathbf{w}_j^a = (\mathbf{u}_j^a, \{\mathbf{v}_j^a\}^c) \quad (9)$$

where

$$\mathbf{u}_j^a = \wedge_{i=1}^P \mathbf{a}(i) \quad (10)$$

and

$$\mathbf{v}_j^a = \vee_{i=1}^P \mathbf{a}(i). \quad (11)$$

Based on the aforementioned expression for $\mathbf{w}_j^a$, we can now state that the weight vector $\mathbf{w}_j^a$ can be expressed in terms of the two $M_a$-dimensional vectors $\mathbf{u}_j^a$ and $\mathbf{v}_j^a$. Hence, the weight vector $\mathbf{w}_j^a$ can be represented, geometrically, in terms of two points in the $M_a$-dimensional space, $\mathbf{u}_j^a$ and $\mathbf{v}_j^a$. Another way of looking at it is that $\mathbf{w}_j^a$ can be represented, geometrically, in terms of a hyperrectangle $R_j^a$ with endpoints $\mathbf{u}_j^a$ and $\mathbf{v}_j^a$ (see Figure 3 for an illustration of this when $M_a = 2$). For simplicity, in this paper we refer to *hyperrectangles* as *rectangles* because most of our illustrations are in two-dimensional space.

Obviously, the aforementioned representation implies that we can geometrically represent an input pattern $\mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$ by a rectangle with endpoints $\mathbf{a}$ and $\mathbf{a}^c$. In other words, $\mathbf{I}$ can be represented by a rectangle of size 0, which is the single point $\mathbf{a}$ in the $M_a$-dimensional space. Note that the *size of a rectangle* $R_j^a$ with endpoints $\mathbf{u}_j^a$ and $\mathbf{v}_j^a$ is taken to be equal to the norm of the vector $\mathbf{v}_j^a - \mathbf{u}_j^a$. The *norm of a vector* in Fuzzy ART is defined to be equal to the sum of the absolute values of its components.

In summary, we will treat $\mathbf{w}_j^a = (\mathbf{u}_j^a, \{\mathbf{v}_j^a\}^c)$ as a rectangle $R_j^a$ with endpoints $\mathbf{u}_j^a$ and $\mathbf{v}_j^a$ in $M_a$-dimensional space, and $\mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$ as the point $\mathbf{a}$ in $M_a$-dimensional space.

The reason why the rectangle representation of a template $\mathbf{w}_j^a$ is so useful is explained below. Consider the template $\mathbf{w}_j^{a,old}$, and its geometrical representative, the rectangle $R_j^{a,old}$ with endpoints $\mathbf{w}_j^{a,old}$ and $\mathbf{v}_j^{a,old}$. Assume that $\mathbf{u}_j^{a,old} = \wedge_{i=1}^P \mathbf{a}(i)$ and $\mathbf{v}_j^{a,old} = \vee_{i=1}^P \mathbf{a}(i)$. Let us now present pattern $\hat{\mathbf{I}} = (\hat{\mathbf{a}}, \hat{\mathbf{a}}^c)$ to Fuzzy ART. Recall that the quantities defined above with a superscript {a, old} indicate values of these quantities prior to the presentation of $\hat{\mathbf{I}}$ to Fuzzy ART. Suppose that, during $\hat{\mathbf{I}}$'s presentation to Fuzzy ART, node $j$ in the $F_2^a$ field is chosen and node $j$ with corresponding weight vector $\mathbf{w}_j^{a,old}$ is appropriate to represent the input pattern $\hat{\mathbf{I}}$. We now distinguish two cases.

In case 1 we assume that $\hat{\mathbf{I}}$ lies inside the rectangle $R_j^{a,old}$ that geometrically represents the template $\mathbf{w}_j^{a,old}$ (see Figure 4). According to the Fuzzy ART rules

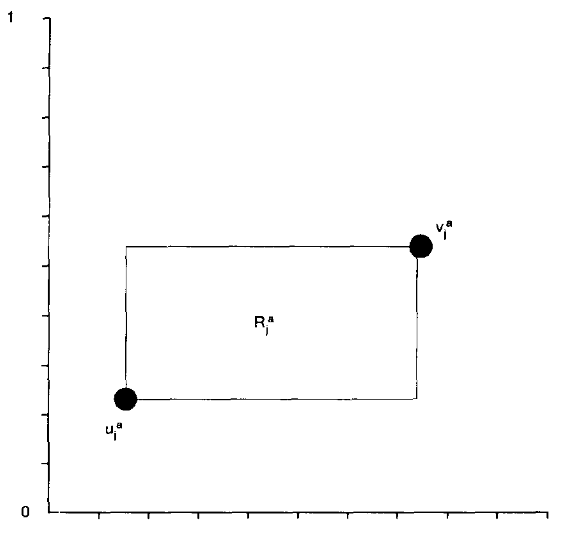


**FIGURE 3. Representation of the template $\mathbf{w}_j^a = \left(\mathbf{u}_j^a, \left\{\mathbf{v}_j^a\right\}^c\right)$ in terms of the rectangle $R_j^a$ with endpoints $\mathbf{u}_j^a$ and $\mathbf{v}_j^a$ (in the figure $M_a = 2$).**

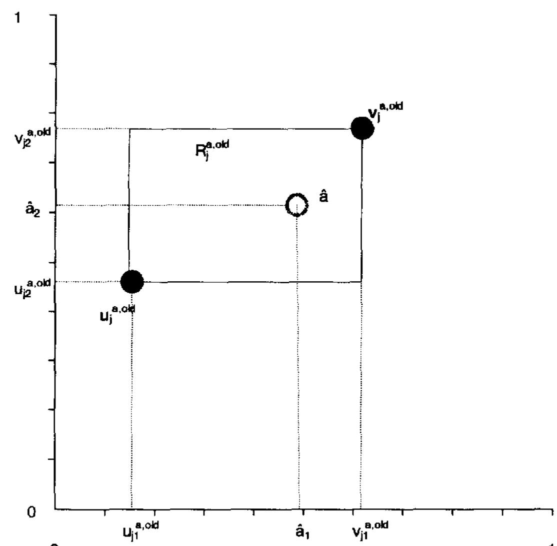


**FIGURE 4. Input pattern $\hat{\mathbf{I}} = (\hat{\mathbf{a}}, \hat{\mathbf{a}}^c)$, represented by the point $\hat{\mathbf{a}}$, lies inside rectangle $R_j^{a,old}$ that represents template $\mathbf{w}_j^{a,old} = (\mathbf{u}_j^{a,old}, \{\mathbf{v}_j^{a,old}\}^c)$. Learning of $\hat{\mathbf{I}}$ leaves $R_j^{a,old}$ intact.**

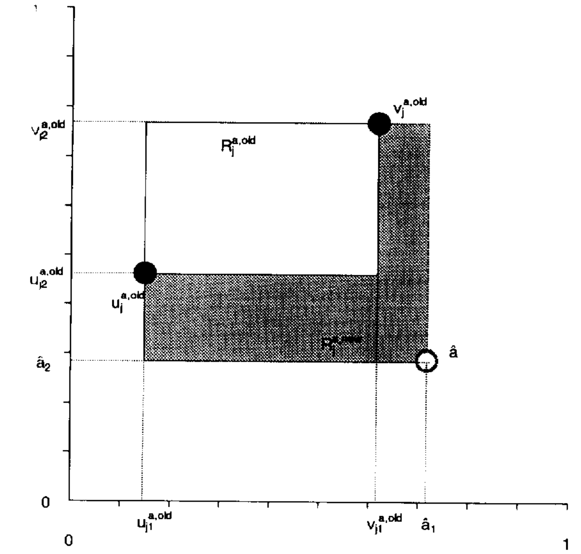


**FIGURE 5. Input pattern $\hat{\mathbf{I}} = (\hat{\mathbf{a}}, \hat{\mathbf{a}}^c)$, represented by the point $\hat{\mathbf{a}}$, lies outside rectangle $R_j^{a,old}$ that represents template $\mathbf{w}_j^{a,old} = (\mathbf{u}_j^{a,old}, \{\mathbf{v}_j^{a,old}\}^c)$. Learning of $\hat{\mathbf{I}}$ creates a new rectangle $R_j^{a,new}$ (the rectangle including all the points of rectangle $R_j^{a,old}$ and the point $\hat{\mathbf{a}}$) of larger size than $R_j^{a,old}$.**

$\mathbf{w}_j^{a,old}$ now becomes equal to $\mathbf{w}_j^{a,new}$, where

$$\begin{aligned}\mathbf{w}_j^{a,new} &= \mathbf{w}_j^{a,old} \wedge \hat{\mathbf{I}} = ((\mathbf{u}_j^{a,old} \wedge \hat{\mathbf{a}}), \{\mathbf{v}_j^{a,old} \vee \hat{\mathbf{a}}\}^c) \\ &= (\mathbf{u}_j^{a,old}, \{\mathbf{v}_j^{a,old}\}^c) = \mathbf{w}_j^{a,old}.\end{aligned}$$

In this case there is no actual weight change, or equivalently, the size of the rectangle that represents the template $\mathbf{w}_j^{a,old}$ remains unchanged.

In case 2, we assume that $\hat{\mathbf{I}}$ lies outside the rectangle $R_j^{a,old}$ that geometrically represents template $\mathbf{w}_j^{a,old}$ (see Figure 5). Once more, according to the Fuzzy ART rules, $\mathbf{w}_j^{a,old}$ becomes equal to $\mathbf{w}_j^{a,new}$, where

$$\begin{aligned}\mathbf{w}_j^{a,new} &= \mathbf{w}_j^{a,old} \wedge \hat{\mathbf{I}} = (\mathbf{u}_j^{a,old} \wedge \hat{\mathbf{a}}, \{\mathbf{v}_j^{a,old} \vee \hat{\mathbf{a}}\}^c) \\ &= (u_{j1}^{a,old} \wedge \hat{a}, \ldots, u_{jM_a}^{a,old} \wedge \hat{a}_{M_a}, (v_{j1}^{a,old} \vee \hat{a}_1)^c, \ldots \\ &\quad (v_{jM_a}^{a,old} \vee \hat{a}_{M_a})^c) \\ &\neq (\mathbf{u}_j^{a,old}, \{\mathbf{v}_j^{a,old}\}^c) = \mathbf{w}_j^{a,old}.\end{aligned} \tag{12}$$

In this case there is actual weight change; the size of the rectangle that is defined by the new weight vector $\mathbf{w}_j^{a,new}$ is increased. Thus, during the training process of Fuzzy ART the size of a rectangle $R_j^a$, that the weight vector $\mathbf{w}_j^a$ defines, can only increase from the size of zero of possibly a maximum size which will be determined below.

The maximum size of a rectangle is determined by the vigilance parameter $\rho_a$. More specifically, with complement coding the size of an input pattern $\mathbf{I}$ is equal to $M_a$. Hence, a node $j$ in the $F_2^a$ field with corresponding weight vector $\mathbf{w}_j^{a,old}$ codes an input pattern $\mathbf{I}$ if the following criterion is satisfied:

$$\left|\mathbf{I} \wedge \mathbf{w}_j^{a,old}\right| \geqslant M_a \rho_a. \tag{13}$$

However,

$$\begin{aligned}|\mathbf{I} \wedge \mathbf{w}_j^{a,old}| &= |(\mathbf{a}, \mathbf{a}^c) \wedge (\mathbf{u}_j^{a,old}, \{\mathbf{v}_j^{a,old}\}^c)| \\ &= |(\mathbf{a} \wedge \mathbf{u}_j^{a,old}, \mathbf{a}^c \wedge \{\mathbf{v}_j^{a,old}\}^c)| \\ &= |(\mathbf{a} \wedge \mathbf{u}_j^{a,old}, (\mathbf{a} \vee \mathbf{v}_j^{a,old})^c)| \\ &= \sum_{i=1}^{M_a} (a_i \wedge u_{ji}^{a,old}) + \sum_{i=1}^{M_a} (a_i \vee v_{ji}^{a,old})^c \\ &= \sum_{i=1}^{M_a} (a_i \wedge u_{ji}^{a,old}) + M_a - \sum_{i=1}^{M_a} (a_i \vee v_{ji}^{a,old}) \\ &= M_a - |(\mathbf{a} \vee \mathbf{v}_j^{a,old}) - (\mathbf{a} \wedge \mathbf{u}_j^{a,old})| \\ &= M_a - |R_j^{a,new}|.\end{aligned} \tag{14}$$

From the above equations we can see that the rectangle size is allowed to increase provided that the new rectangle size satisfies the constraint

$$|R_j^{a,new}| \leqslant M_a(1 - \rho_a).$$

The above inequality implies that if we choose $\rho_a$ small (i.e., $\rho_a \approx 0$), then some of the rectangles that the Fuzzy ART architecture defines might fill most of the entire input pattern space. On the other hand, if $\rho_a$ is close to 1, all of the rectangles will be small.

### 2.5. The Definition of Distance in Fuzzy ART

For the results that are reported in this paper it is important to define the distance of an input pattern $\mathbf{I}$ from a rectangle $R^a$ that does not contain $\mathbf{I}$. We have already verified that every top-down weight vector $\mathbf{w}^a$ in Fuzzy ART has a geometrical interpretation, in $M_a$-dimensional space, in terms of a rectangle with endpoints $\mathbf{u}^a$ and $\mathbf{v}^a$ (see for example Figure 6a, where $M_a = 2$). Also an input pattern $\mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$ can be geometrically represented by the $M_a$-dimensional input vector $\mathbf{a}$ (see for example Figure 6a). The distance between two points $\mathbf{x} = (x_1, \ldots, x_{M_a})$ and $\mathbf{y} = (y_1, \ldots, y_{M_a})$ in $M_a$-dimensional space is defined by

$$\text{dis}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{M_a} |x_i - y_i|. \tag{15}$$

Based on the above definition we can now define the distance of an input pattern $\mathbf{I}$ from a rectangle $R^a$ that does not contain $\mathbf{I}$ (see Figures 6a–6b) as follows.

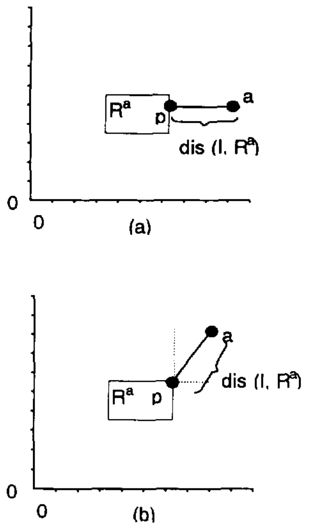


**FIGURE 6. Distance of input pattern $\mathbf{I} = (\hat{\mathbf{a}}, \hat{\mathbf{a}}^c)$, represented by the point a, from the rectangle $R^a$ for different locations of a with respect to the rectangle $R^a$ (a is outside rectangle $R^a$).**

DEFINITION. *The distance of an input pattern $\mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$ from a rectangle $R^a$ that does not contain* **I**, *denoted by*

$$\mathrm{dis}(\mathbf{I}, R^a) \tag{16}$$

*is defined to be the distance of* **a** *from the point* **p** *on the boundary $B_{R^a}$ of the rectangle $R^a$ that minimizes this distance.*

The distance of the input pattern **I** from the rectangle $R^a$ is shown in Figures 6a–6b for various respective locations of **a** with respect to the rectangle $R^a$.

It is not difficult to show that for an input pattern **I** outside a rectangle $R^{a,old}$ the following equality is true:

$$|R^{a,new}| = |R^{a,old}| + \mathrm{dis}(\mathbf{I}, R^{a,old}) \tag{17}$$

where $R^{a,new}$ is the new rectangle created by the input pattern **I** and the old rectangle $R^{a,old}$, if the input pattern **I** chooses and is coded by the node whose weight is represented by rectangle $R^{a,old}$.

## 3. RESULTS

In this section, we report results pertaining to the order in which categories are chosen in the $F_2^a$ field of Fuzzy ART for various ranges of the choice parameter values: $\alpha_a$ small, $\alpha_a$ large, and $\alpha_a$ of intermediate value. The effect on the order of choices made by Fuzzy ART, due to the other two Fuzzy ART parameters ($M_u^a$ and $\rho_a$), is implicit, and as a result, more difficult to investigate. The implicit effect that $M_u^a$ and $\rho_a$ have on the order of choices that Fuzzy ART makes is that $M_u^a$ and $\rho_a$ create coarse or fine clusters (large or small rectangles), depending on their actual values.

Before presenting our main results (Results A, B, and C) we state and prove three theorems that are used to demonstrate the results. Results A, B, and C describe the order of choices made by Fuzzy ART for $\alpha_a$ small, $\alpha_a$ large, and intermediate $\alpha_a$ values, respectively. Because Fuzzy ART exhibits a distinctly different behavior for different values of $\alpha_a$, presenting the results in this fashion gives us a better understanding of the order of choices made by Fuzzy ART.

THEOREM 1. *If an input pattern* **I** *is presented to Fuzzy ART, and* **I** *is inside rectangles $R_{j_1}^{a,old}$ and $R_{j_2}^{a,old}$, then* **I** *will first choose the rectangle of the smallest size.*

*Proof.* Assume that rectangle $R_{j_1}^{a,old}$ is smaller than rectangle $R_{j_2}^{a,old}$. That is,

$$|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}|. \tag{18}$$

The bottom-up input to node $j_1$ due to the presentation of pattern **I** is equal to:

$$T_{j_1}^a = \frac{|\mathbf{I} \wedge \mathbf{w}_{j_1}^{a,old}|}{\alpha_a + |\mathbf{w}_{j_1}^{a,old}|} = \frac{|\mathbf{w}_{j_1}^{a,old}|}{\alpha_a + |\mathbf{w}_{j_1}^{a,old}|} = \frac{M_a - |R_{j_1}^{a,old}|}{\alpha_a + M_a - |R_{j_1}^{a,old}|}. \tag{19}$$

The bottom-up input to node $j_2$ due to the presentation of pattern **I** is equal to:

$$T_{j_2}^a = \frac{|\mathbf{I} \wedge \mathbf{w}_{j_2}^{a,old}|}{\alpha_a + |\mathbf{w}_{j_2}^{a,old}|} = \frac{|\mathbf{w}_{j_2}^{a,old}|}{\alpha_a + |\mathbf{w}_{j_2}^{a,old}|} = \frac{M_a - |R_{j_2}^{a,old}|}{\alpha_a + M_a - |R_{j_2}^{a,old}|}. \tag{20}$$

Due to eqn (18) we can deduce that

$$M_a - |R_{j_1}^{a,old}| > M_a - |R_{j_2}^{a,old}|. \tag{21}$$

Hence, from eqns (19), (20) and (21) we conclude that

$$T_{j_1}^a > T_{j_2}^a \tag{22}$$

and, according to the Fuzzy ART rules, rectangle $R_{j_1}^{a,old}$ will be chosen prior to rectangle $R_{j_2}^{a,old}$.

THEOREM 2. *If an input pattern* **I** *is presented to Fuzzy ART, and* **I** *is outside rectangle $R_{j_1}^{a,old}$ and inside rectangle $R_{j_2}^{a,old}$, then* **I** *will first choose rectangle $R_{j_1}^{a,old}$ iff*

$$\mathrm{dis}(\mathbf{I}, R_{j_1}^{a,old}) < g(\alpha_a) \tag{23}$$

where

$$g(\alpha_a) = \left[\frac{M_a - |R_{j_1}^{a,old}|}{\alpha_a + M_a - |R_{j_1}^{a,old}|} - \frac{M_a - |R_{j_2}^{a,old}|}{\alpha_a + M_a - |R_{j_2}^{a,old}|}\right] \times (\alpha_a + M_a - |R_{j_1}^{a,old}|). \quad (24)$$

*Proof.* The bottom-up input to node $j_1$ due to the presentation of pattern **I** is equal to:

$$T_{j_1}^a = \frac{|\mathbf{I} \wedge \mathbf{w}_{j_1}^{a,old}|}{\alpha_a + |\mathbf{w}_{j_1}^{a,old}|} = \frac{M_a - |R_{j_1}^{a,new}|}{\alpha_a + M_a - |R_{j_1}^{a,old}|}. \quad (25)$$

The bottom-up input to node $j_2$ due to the presentation of pattern **I** is equal to:

$$T_{j_2}^a = \frac{|\mathbf{I} \wedge \mathbf{w}_{j_2}^{a,old}|}{\alpha_a + |\mathbf{w}_{j_2}^{a,old}|} = \frac{|\mathbf{w}_{j_2}^a|}{\alpha_a + |\mathbf{w}_{j_2}^a|} = \frac{M_a - |R_{j_2}^{a,old}|}{\alpha_a + M_a - |R_{j_2}^{a,old}|}. \quad (26)$$

Suppose now that

$$T_{j_1}^a > T_{j_2}^a. \quad (27)$$

Then, substituting $T_{j_1}^a$ and $T_{j_2}^a$ from eqns (25) and (26) into inequality (27) we get

$$\frac{M_a - |R_{j_1}^{a,new}|}{\alpha_a + M_a - |R_{j_1}^{a,old}|} > \frac{M_a - |R_{j_2}^{a,old}|}{\alpha_a + M_a - |R_{j_2}^{a,old}|}. \quad (28)$$

If in the above inequality we substitute $|R_{j_1}^{a,new}|$ with its equal from below

$$|R_{j_1}^{a,new}| = |R_{j_1}^{a,old}| + dis(\mathbf{I}, R_{j_1}^{a,old}) \quad (29)$$

we derive, after minor manipulations, inequality (23). Hence, inequality (23) is a necessary and sufficient condition for pattern **I** to choose rectangle $R_{j_1}^{a,old}$ prior to choosing rectangle $R_{j_2}^{a,old}$ (provided that the assumptions of the theorem are valid).

THEOREM 3. *If an input pattern* **I** *is presented to Fuzzy ART, and* **I** *is outside rectangles* $R_{j_1}^{a,old}$ *and* $R_{j_2}^{a,old}$, *then* **I** *will first choose* $R_{j_1}^{a,old}$ *iff*

$$\text{dis}(\mathbf{I}, R_{j_1}^{a,old}) < h(\alpha_a) \quad (30)$$

where

$$h(\alpha_a) = \frac{\alpha_a + M_a - |R_{j_1}^{a,old}|}{\alpha_a + M_a - |R_{j_2}^{a,old}|} dis(\mathbf{I}, R_{j_2}^{a,old}) + \frac{\alpha_a}{\alpha_a + M_a - |R_{j_2}^{a,old}|}(|R_{j_2}^{a,old}| - |R_{j_1}^{a,old}|). \quad (31)$$

*Proof.* The bottom-up input to node $j_1$ due to the presentation of pattern **I** is equal to:

$$T_{j_1}^a = \frac{|\mathbf{I} \wedge \mathbf{w}_{j_1}^{a,old}|}{\alpha_a + |\mathbf{w}_{j_1}^{a,old}|} = \frac{M_a - |R_{j_1}^{a,new}|}{\alpha_a + M_a - |R_{j_1}^{a,old}|}. \quad (32)$$

The bottom-up input to node $j_2$ due to the presentation of pattern **I** is equal to:

$$T_{j_2}^a = \frac{|\mathbf{I} \wedge \mathbf{w}_{j_2}^{a,old}|}{\alpha_a + |\mathbf{w}_{j_2}^{a,old}|} = \frac{M_a - |R_{j_2}^{a,new}|}{\alpha_a + M_a - |R_{j_2}^{a,old}|}. \quad (33)$$

Suppose now that

$$T_{j_1}^a > T_{j_2}^a. \quad (34)$$

Then, substituting $T_{j_1}^a$ and $T_{j_2}^a$ from eqns (32) and (33) into inequality (34) we get

$$\frac{M_a - |R_{j_1}^{a,new}|}{\alpha_a + M_a - |R_{j_1}^{a,old}|} > \frac{M_a - |R_{j_2}^{a,new}|}{\alpha_a + M_a - |R_{j_2}^{a,old}|}. \quad (35)$$

If, in the above inequality, we substitute $R_{j_1}^{a,new}$ and $R_{j_2}^{a,new}$ with their equals from below

$$|R_{j_1}^{a,new}| = |R_{j_1}^{a,old}| + dis(\mathbf{I}, R_{j_1}^{a,old}) \quad (36)$$

and

$$|R_{j_2}^{a,new}| = |R_{j_2}^{a,old}| + dis(\mathbf{I}, R_{j_2}^{a,old}) \quad (37)$$

we derive, after minor manipulations, that rectangle $R_{j_1}^{a,old}$ will be chosen first if the following inequality is satisfied.

$$\text{dis}(\mathbf{I}, R_{j_1}^{a,old}) < \frac{\alpha_a + M_a - |R_{j_1}^{a,old}|}{\alpha_a + M_a - |R_{j_2}^{a,old}|}\text{dis}(\mathbf{I}, R_{j_2}^{a,old}) + (M_a - |R_{j_1}^{a,old}|) - \frac{\alpha_a + M_a - |R_{j_1}^{a,old}|}{\alpha_a + M_a - |R_{j_2}^{a,old}|}(M_a - |R_{j_2}^{a,old}|). \quad (38)$$

If we combine the last two terms in the right hand side of the above inequality we end up with inequality (30), being the necessary and sufficient condition for the input pattern **I** to choose rectangle $R_{j_1}^{a,old}$ prior to choosing rectangle $R_{j_2}^{a,old}$ (provided of course that the assumptions of the theorem are valid).

Theorems 1–3 are now used to derive the main results of the paper.

RESULT A. *If an input pattern* **I** *is presented to a Fuzzy ART architecture with small* $\alpha_a$ *parameter values (i.e.,* $\alpha_a$ *close to zero), and*

1. **I** *is inside rectangles* $R_{j_1}^{a,old}$ *and* $R_{j_2}^{a,old}$, *then* **I** *will first choose the rectangle of the smallest size.*
2. **I** *is outside rectangle* $R_{j_1}^{a,old}$ *and inside rectangle* $R_{j_2}^{a,old}$, *then* **I** *will first choose rectangle* $R_{j_2}^{a,old}$.
3. **I** *is outside rectangles* $R_{j_1}^{a,old}$ *and* $R_{j_2}^{a,old}$, *then* **I** *will first choose rectangle* $R_{j_1}^{a,old}$ *iff*

$$\mathrm{dis}(\mathbf{I}, R_{j_1}^{a,old}) < \frac{M_a - |R_{j_1}^{a,old}|}{M_a - R_{j_2}^{a,old}|}\mathrm{dis}(\mathbf{I}, R_{j_2}^{a,old}). \tag{39}$$

*Proof. Result A1.* Result A1 is actually a special case ($\alpha_a \approx 0$) of Theorem 1.

*Result A2.* Result A2 is an immediate consequence of Theorem 2. Theorem 2 says that rectangle $R_{j_1}^{a,old}$ will be chosen first iff

$$\mathrm{dis}(\mathbf{I}, R_{j_1}^{a,old}) < g(\alpha_a). \tag{40}$$

But $g(\alpha_a) \approx 0$ when $\alpha_a \approx 0$. Hence, the above inequality cannot happen if $\alpha_a$ is approximately equal to 0. Hence, **I** will always choose first rectangle $R_{j_2}^{a,old}$.

*Result A3.* Result A3 is a direct result of Theorem 3. Theorem 3 stated that if input pattern **I** is outside rectangles $R_{j_1}^{a,old}$ and $R_{j_2}^{a,old}$, then **I** will first choose $R_{j_1}^{a,old}$ iff

$$\mathrm{dis}(\mathbf{I}, R_{j_1}^{a,old}) < h(\alpha_a). \tag{41}$$

But since $\alpha_a \approx 0$, $h(\alpha_a) \approx h(0)$, where

$$h(0) = \frac{M_a - |R_{j_1}^{a,old}|}{M_a - |R_{j_2}^{a,old}|} dis(\mathbf{I}, R_{j_2}^{a,old}). \tag{42}$$

The above arguments verify Result A3.

*Result B. If an input pattern* **I** *is presented to a Fuzzy ART architecture with large* $\alpha_a$ *parameter values (i.e.,* $\alpha_a$ *approaching* $\infty$*), and*

1. **I** *is inside rectangles* $R_{j_1}^{a,old}$ *and* $R_{j_2}^{a,old}$, *then* **I** *will first choose the rectangle of the smallest size.*
2. **I** *is outside rectangle* $R_{j_1}^{a,old}$ *and inside rectangle* $R_{1}^{a,old}$, *then* **I** *will first choose rectangle* $R_{j_1}^{a,old}$ *iff*

$$\mathrm{dis}(\mathbf{I}, R_{j_1}^{a,old}) < |R_{j_2}^{a,old}| - |R_{j_1}^{a,old}| \tag{43}$$

*or equivalently iff*

$$|R_{j_1}^{a,new}| < |R_{j_2}^{a,old}|. \tag{44}$$

3. **I** *is outside rectangles* $R_{j_1}^{a,old}$ *and* $R_{j_2}^{a,old}$, *then* **I** *will first choose rectangle* $R_{j_1}^{a,old}$ *iff*

$$dis(\mathbf{I}, R_{j_1}^{a,old}) < dis(\mathbf{I}, R_{j_2}^{a,old}) + |R_{j_2}^{a,old}| - |R_{j_1}^{a,old}| \tag{45}$$

*or equivalently iff*

$$|R_{j_1}^{a,new}| < |R_{j_2}^{a,new}|. \tag{46}$$

*Proof. Result B1.* Result B1 is actually a special case ($\alpha_a$ approaches $\infty$) of Theorem 1.

*Result B2.* Result B2 is an immediate consequence of Theorem 2. Theorem 2 sates that rectangle $R_{j_1}^{a,old}$ will be chosen first iff

$$\mathrm{dis}(\mathbf{I}, R_{j_1}^{a,old}) < g(\alpha_a). \tag{47}$$

But when $\alpha_a \to \infty$, $g(\alpha_a)$ approaches

$$|R_{j_2}^{a,old}| - |R_{j_1}^{a,old}|. \tag{48}$$

Hence, the validity of Result B2 is evident.

*Result B3.* Result B3 is a direct result of Theorem 3. Theorem 3 states that if the input pattern **I** is outside rectangles $R_{j_1}^{a,old}$ and $R_{j_2}^{a,old}$, then **I** will first choose first rectangle $R_{j_1}^{a,old}$ iff

$$\mathrm{dis}(\mathbf{I}, R_{j_1}^{a,old}) < h(\alpha_a). \tag{49}$$

But since $\alpha_a \to \infty$, $h(\alpha_a) \to h(\infty)$, where

$$h(\infty) = \mathrm{dis}(\mathbf{I}, R_{j_2}^{a,old}) + |R_{j_2}^{a,old}| - |R_{j_1}^{a,old}|. \tag{50}$$

The above equation verifies the validity of Result B3.

*Result C. If an input pattern* **I** *is presented to a Fuzzy ART architecture with intermediate* $\alpha_a$ *parameter values (i.e.,* $0 < \alpha_a < \infty$*), and*

1. **I** *is inside rectangles* $R_{j_1}^{a,old}$ *and* $R_{j_2}^{a,old}$, *then* **I** *will first choose the rectangle of the smallest size.*
2. **I** *is outside rectangle* $R_{j_1}^{a,old}$ *and inside rectangle* $R_{j_2}^{a,old}$, *then* **I** *will first choose rectangle* $R_{j_1}^{a,old}$ *iff*

$$\mathrm{dis}(\mathbf{I}, R_{j_1}^{a,old}) < g(\alpha_a) \tag{51}$$

*where, given that*

$$|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}|, \quad (52)$$

*$g(\alpha_a)$ is a nondecreasing function of $\alpha_a$, and*

$$0 < g(\alpha_a) < -|R_{j_2}^{a,old}| - |R_{j_1}^{a,old}|. \quad (53)$$

3. **I** *is outside rectangles $R_{j_1}^{a,old}$ and $R_{j_2}^{a,old}$, then* **I** *will first choose rectangle $R_{j_1}^{a,old}$ iff*

$$\text{dis}(\mathbf{I}, R_{j_1}^{a,old}) < h(\alpha_a) \quad (54)$$

*where, given that*

$$|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}|, \quad (55)$$

*$h(\alpha_a)$ is a nondecreasing function of $\alpha_a$, and*

$$\frac{M_a - |R_{j_1}^{a,old}|}{M_a - |R_{j_2}^{a,old}|}\text{dis}(\mathbf{I}, R_{j_2}^{a,old}) < h(\alpha_a) < \text{dis}(\mathbf{I}, R_{j_2}^{a,old}) + |R_{j_2}^{a,old}| - |R_{j_1}^{a,old}|. \quad (56)$$

*Proof. Result C1.* Result C1 is a special case of Theorem 1.

*Result C2.* The first statement of Result C2 is actually Theorem 2, where $0 < \alpha_a < \infty$. To demonstrate the second statement of Result C2, consider the function $g(\alpha_a)$ defined in the statement of Theorem 2. With a little algebraic manipulation it is easy to show that

$$\frac{dg(\alpha_a)}{d\alpha_a} = -\frac{(M_a - |R_{j_2}^{a,old}|)(|R_{j_1}^{a,old}| - |R_{j_2}^{a,old}|)}{(\alpha_a + M_a - |R_{j_2}^{a,old}|)^2}. \quad (57)$$

The following equalities are true:

$$M_a - |R_{j_2}^{a,old}| \geqslant 0 \quad (58)$$

and

$$|R_{j_1}^{a,old}| - |R_{j_2}^{a,old}| < 0. \quad (59)$$

Inequality (58) is obvious, while inequality (59) is an assumption of Result C2. Utilizing the above inequalities in eqn (57) we obtain

$$\frac{dg(\alpha_\alpha)}{d\alpha_a} \geqslant 0, \quad (60)$$

which implies that $g(\alpha_\alpha)$ is a nondecreasing function of its argument $\alpha_a$. Consequently,

$$\inf_{\alpha_a} g(\alpha_a) = \lim_{\alpha_a \to 0} g(\alpha_a) = g(0) = 0 \quad (61)$$

and

$$\sup_{\alpha_a} g(\alpha_a) = \lim_{\alpha_a \to 0} g(\alpha_a) = |R_{j_2}^{a,old}| - |R_{j_1}^{a,old}|. \quad (62)$$

*Result C3.* The first statement of Result C3 is actually Theorem 3. The statement about the behavior of $h(\alpha_\alpha)$ can be proven by considering the derivative of $h(\alpha_\alpha)$. With a little algebraic manipulation it can be shown that the derivative of $h(\alpha_\alpha)$ with respect to $\alpha_a$ is equal to:

$$\frac{dh(\alpha_a)}{d\alpha_a} = \frac{(M_a - |R_{j_2}^{a,new}|)(|R_{j_2}^{a,old}| - |R_{j_1}^{a,old}|)}{(\alpha_a + M_a - |R_{j_2}^{a,old}|)^2}. \quad (63)$$

Since,

$$|R_{j_2}^{a,old}| > |R_{j_1}^{a,old}| \quad (64)$$

and

$$|R_{j_2}^{a,new}| \leqslant M_a, \quad (65)$$

we conclude from eqn (63) that $h(\alpha_a)$ is a nondecreasing function of its argument $\alpha_a$. Thus,

$$\inf_{\alpha_a} h(\alpha_a) = \lim_{\alpha_a \to 0} h(\alpha_a) = h(0) = \frac{M_a - |R_{j_1}^{a,old}|}{M_a - |R_{j_2}^{a,old}|} dis(\mathbf{I}, R_{j_2}^{a,old}) \quad (66)$$

and

$$\sup_{\alpha_a} h(\alpha_a) = \lim_{\alpha_a \to \infty} h(\alpha_a) = \text{dis}(\mathbf{I}, R_{j_2}^{a,old}) + |R_{j_2}^{a,old}| - |R_{j_1}^{a,old}|. \quad (67)$$

## 4. DISCUSSION OF THE RESULTS

Let us now examine Results A, B, and C to establish some order of search rules according to which categories (or their corresponding rectangles) are chosen in the $F_2^a$ field of Fuzzy ART. Orders of search rules are named after the results that generate them (e.g., the order of search rule B2a is the first order of search rule pertaining to Result B2, while order of search rule BC2b is the second order of search rule pertaining to results B2 and C2).

Order of Search Rule ABC1. *If input pattern* **I** *is inside rectangles $R_{j_1}^{a,old}$ and $R_{j_2}^{a,old}$, and $|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}|$, then pattern* **I** *chooses first the rectangle of the smallest size $R_{j_1}^{a,old}$.*

This order of search rule is a restatement of

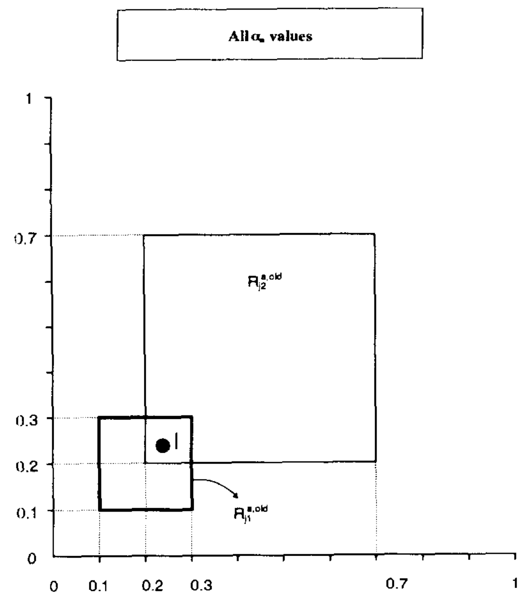


**FIGURE 7. Illustration of the order of search rule ABC1. Pattern I, represented by the point in the figure, is inside rectangles $R_{j_1}^{a,old}$ and $R_{j_2}^{a,old}$. Pattern I chooses first the rectangle of the smallest size $R_{j_1}^{a,old}$. The rectangle of choice is shown in the figure with a bold-faced perimeter.**

Results A1, B1 and C1, and is illustrated in Figure 7 for $M_a = 2$.

ORDER OF SEARCH RULE A2. *If input pattern* **I** *is outside rectangle $R_{j_1}^{a,old}$ and inside rectangle $R_{j_2}^{a,old}$, and $\alpha_a$ is small, then pattern* **I** *chooses first rectangle $R_{j_2}^{a,old}$.*

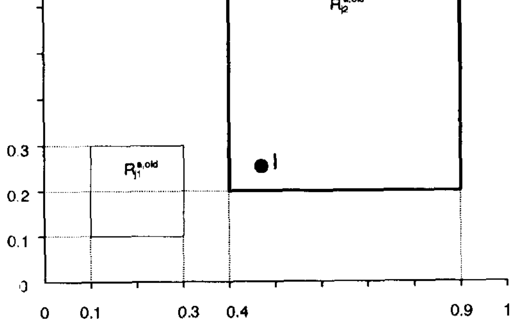


**FIGURE 8. Illustration of the order of search rule A2. Pattern I, represented by the point in the figure, is outside rectangle $R_{j_1}^{a,old}$ and inside rectangle $R_{j_2}^{a,old}$, and $\alpha_a$ is small. Pattern I chooses first rectangle $R_{j_2}^{a,old}$. The rectangle of choice is shown in the figure with a bold-faced perimeter.**

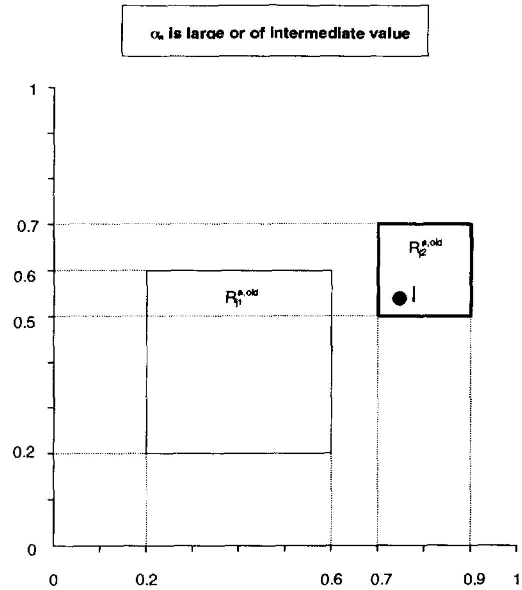


**FIGURE 9. Illustration of the order of search rule BC2a. Pattern I, represented by the point in the figure, is outside rectangle $R_{j_1}^{a,old}$ and inside rectangle $R_{j_2}^{a,old}$, $|R_{j_2}^{a,old}| < |R_{j_1}^{a,old}|$, and $\alpha_a$ is either of intermediate value or of large value. Pattern I chooses first rectangle $R_{j_2}^{a,old}$. The rectangle of choice is shown in the figure with a bold-faced perimeter.**

This order of search rule is a restatement of Result A2, and is illustrated in Figure 8 for $M_a = 2$.

ORDER OF SEARCH RULE BC2a. *If input pattern* **I** *is outside rectangle $R_{j_1}^{a,old}$ and inside rectangle $R_{j_2}^{a,old}$, $|R_{j_2}^{a,old}| < |R_{j_1}^{a,old}|$, and $\alpha_a$ is either of intermediate value or of large value, then pattern* **I** *chooses first rectangle $R_{j_2}^{a,old}$.*

This order of search rule is an immediate consequence of Result B2 [see eqn (43)], for $\alpha_a$ large, and Result C2 [see eqn (51)], for $\alpha_a$ of intermediate value. This order of search rule is illustrated in Figure 9 for $M_a = 2$.

ORDER OF SEARCH RULE BC2b. *If input pattern* **I** *is outside rectangle $R_{j_1}^{a,old}$ and inside rectangle $R_{j_2}^{a,old}$, $|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}| < |R_{j_1}^{a,new}|$, and $\alpha_a$ is either of intermediate value or of large value, then pattern* **I** *chooses first rectangle $R_{j_2}^{a,old}$.*

This order of search rule is also an immediate consequence of Result B2 [see eqn (44)], for $\alpha_a$ large, and Result C2 [see eqns (51), (52) and (53)], for $\alpha_a$ of intermediate value. This order of search rule is illustrated in Figure 10 for $M_a = 2$.

ORDER OF SEARCH RULE B2c. *If an input pattern* **I** *is outside rectangle $R_{j_1}^{a,old}$ and inside rectangle $R_{j_2}^{a,old}$, $|R_{j_1}^{a,new}| < |R_{j_2}^{a,old}|$, and $\alpha_a$ is of large value, then pattern* **I** *chooses first rectangle $R_{j_1}^{a,old}$.*

This order of search rule is also an immediate consequence of Result B2 [see eqn (44)], and is illustrated in Figure 11 for $M_a = 2$.

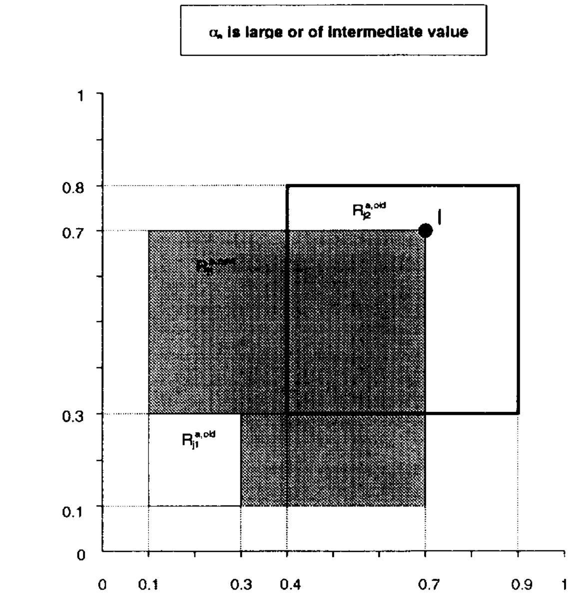


**FIGURE 10. Illustration of the order of search rule BC2b. Pattern I, represented by the point in the figure, is outside rectangle $R_{j_1}^{a,old}$ and inside rectangle $R_{j_2}^{a,old}$, $|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}| < |R_{j_1}^{a,new}|$, and $\alpha_a$ is either of intermediate value or of large value. Pattern I chooses first rectangle $R_{j_2}^{a,old}$. The rectangle of choice is shown in the figure with a bold-faced perimeter.**

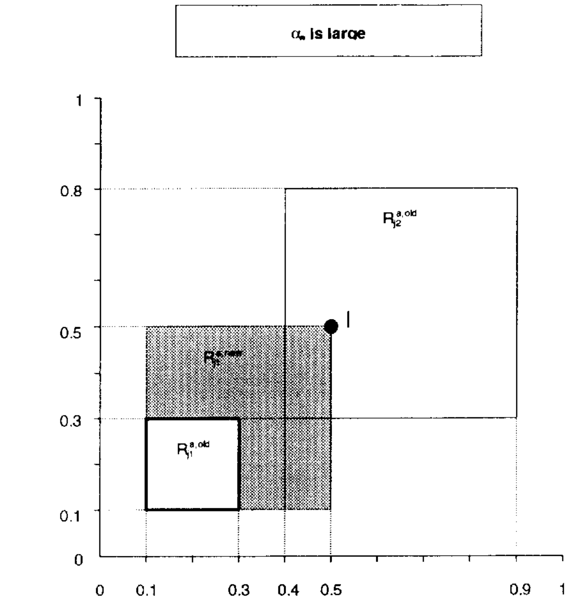


**FIGURE 11. Illustration of the order of search rule B2c. Pattern I, represented by the point in the figure, is outside rectangle $R_{j_1}^{a,old}$ and inside rectangle $R_{j_2}^{a,old}$, $|R_{j_1}^{a,new}| < |R_{j_2}^{a,old}|$, and $\alpha_a$ is a large value. Pattern I chooses first rectangle $R_{j_1}^{a,old}$. The rectangle of choice is shown in the figure with a bold-faced perimeter.**

ORDER OF SEARCH RULE C2c. *If input pattern* **I** *is outside rectangle* $R_{j_1}^{a,old}$ *and inside rectangle* $R_{j_2}^{a,old}$, $|R_{j_1}^{a,new}| < |R_{j_2}^{a,old}|$, *and* $\alpha_a$ *is of intermediate value, then pattern* **I**, *at times, chooses* $R_{j_1}^{a,old}$ *first and, at other times,* $R_{j_2}^{a,old}$ *first. The rectangle chosen depends on the distance of* **I** *from the rectangle that does not contain* **I**, *and the sizes of the rectangles. The frequency with which rectangle* $R_{j_1}^{a,old}$ *is chosen first by a pattern* **I** *does not decrease as* $\alpha_a$ *increases from 0 to* $\infty$.

The first two statements of the above rule are a consequence of Result C2 [see eqns (51), (52), and (53)]. The last statement of the above rule is valid because of eqn (51) and the non-decreasing nature of $g(\alpha_a)$. The above rule is illustrated in Figures 12a–12d for $M_a = 2$.

ORDER OF SEARCH RULE ABC3a. *If pattern* **I** *is outside rectangles* $R_{j_1}^{a,old}$ *and* $R_{j_2}^{a,old}$, $|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}|$, *and* $|R_{j_1}^{a,new}| > |R_{j_2}^{a,new}|$, *then pattern* **I** *first chooses* rectangle $R_{j_2}^{a,old}$.

This order of search rule is an immediate consequence of Results A3, B3 and C3 [see eqns (39), (45), (54), and (56)], and is illustrated in Figure 13 for $M_a = 2$.

ORDER OF SEARCH RULE B3b. *If input pattern* **I** *is outside rectangles* $R_{j_1}^{a,old}$ *and* $R_{j_2}^{a,old}$, $|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}|$, *and* $\alpha_a$ *is large, then pattern* **I** *chooses* $R_{j_1}^{a,old}$ *first if and only if* $|R_{j_1}^{a,new}| < |R_{j_2}^{a,new}|$.

This order of search rule is an immediate consequence of Result B3 [see eqn (45)], and is illustrated in Figures 14a and 14b for $M_a = 2$.

ORDER OF SEARCH RULE C3b. *If input pattern* **I** *is outside rectangles* $R_{j_1}^{a,old}$ *and* $R_{j_2}^{a,old}$, $|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}|$, *and* $\alpha_a$ *is of intermediate value, then pattern* **I**, *at times, chooses* $R_{j_1}^{a,old}$ *first and, at other times, it chooses* $R_{j_2}^{a,old}$ *first. The rectangle chosen depends on the distance of* **I** *from the rectangles and the relative sizes of the rectangles. The frequency with which rectangle* $R_{j_1}^{a,old}$ *is chosen first by a pattern* **I** *does not decrease as* $\alpha_a$ *increases from 0 to* $\infty$.

The first two statements of the above rule are a consequence of Result C3 [see eqns (54), (55), and (56)]. The last statement of the above rule is valid because of eqn (54) and the non-decreasing nature of $h(\alpha_a)$. The above rule is illustrated in Figures 15a–15d for $M_a = 2$.

It is worth pointing out that the results of the previous section (Results A, B, and C) explain how Fuzzy ART chooses among the committed nodes during *off-line training*, or *on-line training*, or the *performance phase*. Results A, B, and C ignore the effect of uncommitted nodes on the Fuzzy ART choices. However, under mild conditions on the Fuzzy ART parameters, we can make the claim that Results A, B, and C describe completely the Fuzzy ART choices *after the first list presentation of an off-line training phase*. To justify this claim let us present a theorem.

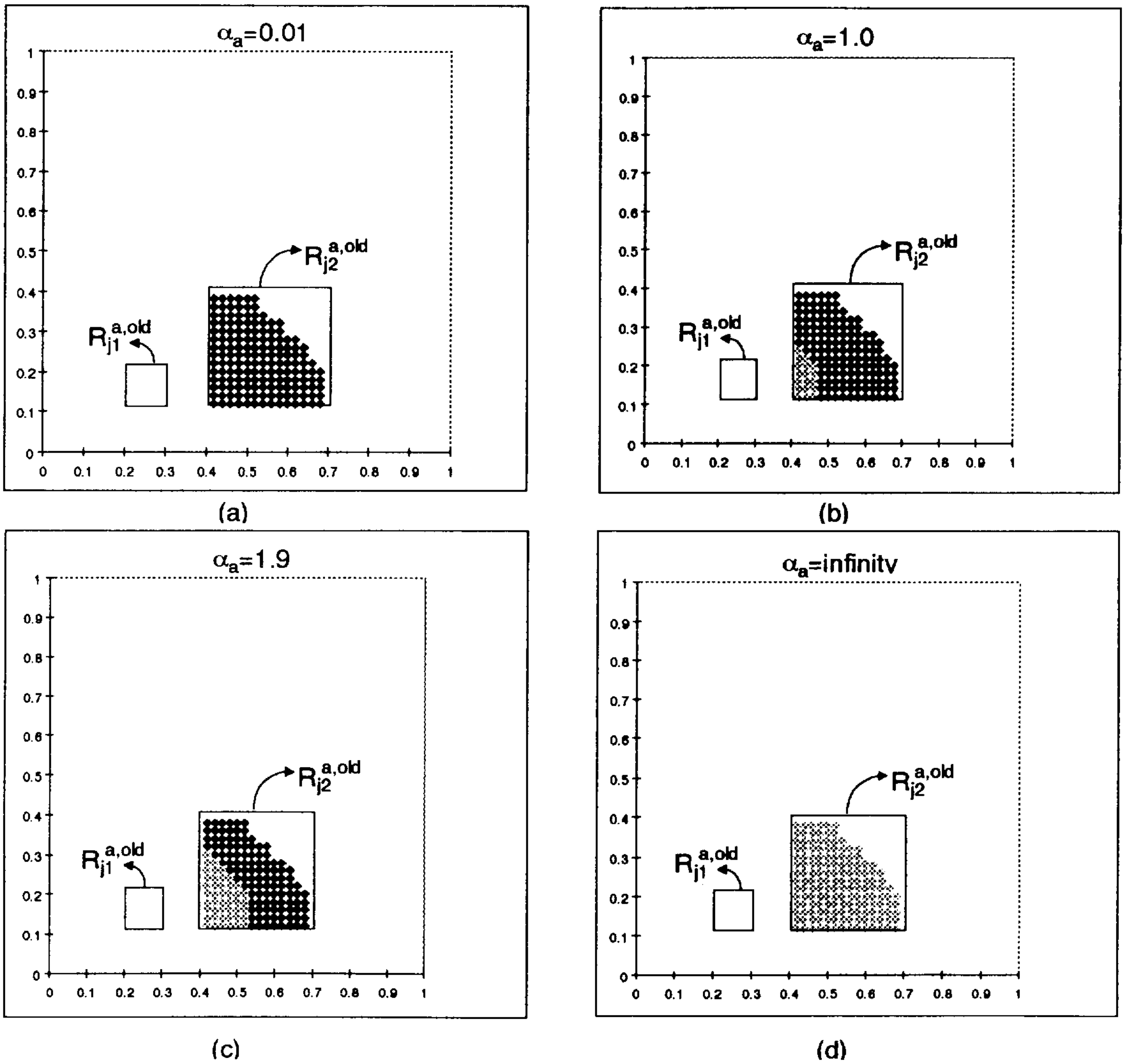


**FIGURE 12 (a,b,c,d). Illustration of the Rule of Thumb C2c. Pattern I is outside rectangle $R_{j_1}^{a,old}$ and inside rectangle $R_{j_2}^{a,old}$, $|R_{j_1}^{a,new}| < |R_{j_2}^{a,old}|$, and $\alpha_\alpha$ is of intermediate value. Pattern I, at times, chooses $R_{j_1}^{a,old}$ first and, at other times, $R_{j_2}^{a,old}$ first. The rectangle chosen depends on the distance of I from the rectangle that does not contain I, and the relative sizes of the rectangles. The frequency with which rectangle $R_{j_1}^{a,old}$ is chosen first by a pattern I does not decrease as $\alpha_\alpha$ increase from 0 to ∞ [compare Figure 12 (a,b,c,d)]. In the figure the gray marks indicate patterns I that choose rectangle $R_{j_1}^{a,old}$ and the black marks indicate patterns I that choose rectangle $R_{j_2}^{a,old}$.**

THEOREM 4. *In Fuzzy ART if*

$$\rho_a > \frac{\alpha_a}{\alpha_a + M_u^a - M_a}, \tag{68}$$

*then uncommitted nodes will not be chosen after the first list presentation of an off-line training phase.*

*Proof.* After the first list presentation of an off-line training phase every input pattern **I** from the training list has a subset template $\mathbf{w}^{a,old}$ that can represent the input pattern (i.e., if $\mathbf{w}^{a,old}$ is chosen it will not be reset).

At this point it is appropriate to define the concept of a subset template in Fuzzy ART. A template $\mathbf{w}^{a,old}$ is a *subset template* of an input pattern **I** if each one of the $\mathbf{w}^{a,old}$ components is smaller than or equal to its corresponding components in **I** (i.e., $w_i^{a,old} \leqslant 1 \leqslant i \leqslant 2M_a$).

A subset template $\mathbf{w}^{a,old}$ will be chosen before an uncommitted node iff

$$\frac{|\mathbf{I}|}{\alpha_a + M_u^a} < \frac{|\mathbf{I} \wedge \mathbf{w}^{a,old}|}{\alpha_a + |\mathbf{w}^{a,old}|} = \frac{|\mathbf{w}^{a,old}|}{\alpha_a + |\mathbf{w}^{a,old}|} = \frac{M_a - |R^{a,old}|}{\alpha_a + M_a - |R^{a,old}|}. \tag{69}$$

However, $M_a - |R^{a,old}| \geqslant M_a\rho_a$ and $|\mathbf{I}| = M_a$. Hence, a sufficient condition for eqn (69) to be valid is

$$\frac{M_a}{\alpha_a + M_u^a} < \frac{M_a\rho_a}{\alpha_a + M_a\rho_a}. \tag{70}$$

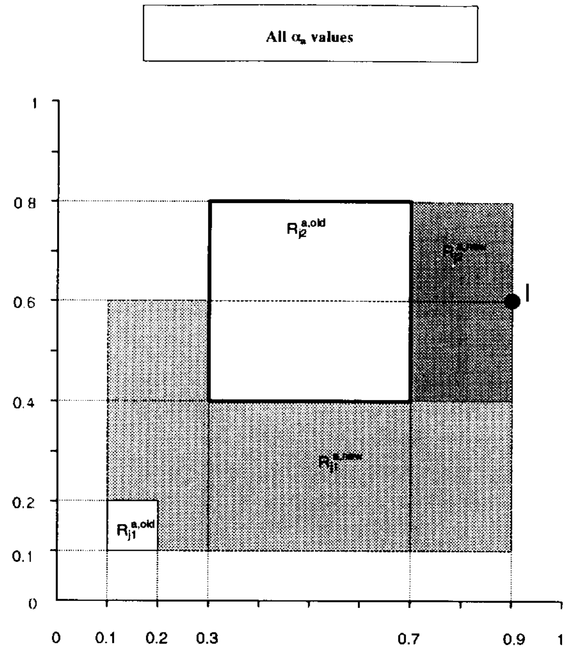


**FIGURE 13 (a,b). Illustration of the order of search rule ABC3a. Pattern I, represented by the point in the figure, is outside rectangles $R_{j_1}^{a,old}$ and $R_{j_2}^{a,old}$, $|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}|$, and $|R_{j_1}^{a,new}| > |R_{j_2}^{a,new}|$. Pattern I first chooses rectangle $R_{j_2}^{a,old}$. The rectangle of choice is shown in the figure with a bold-faced perimeter.**

After minor manipulations, inequality (70) becomes:

$$\rho_a > \frac{\alpha_a}{\alpha_a + M_u^a - M_a}. \quad (71)$$

Hence, under the conditions specified by the above inequality, uncommitted nodes will not be chosen in Fuzzy ART after the first list presentation of an off-line training phase.

It is easy to see that condition (71) is not very stringent, considering that $M_u^a \geqslant 2M_a$ and most Fuzzy ART simulations involve $\alpha_a$ parameter values that are smaller than $M_a$. We are now ready to state another result.

RESULT D. *Results A, B, and C describe completely the order of choices made by the Fuzzy ART architecture after the first list presentation of an off-line training phase, provided that the Fuzzy ART parameters are chosen according to the following rule:*

$$\rho_a > \frac{\alpha_a}{\alpha_a + M_u^a - M_a}. \quad (72)$$

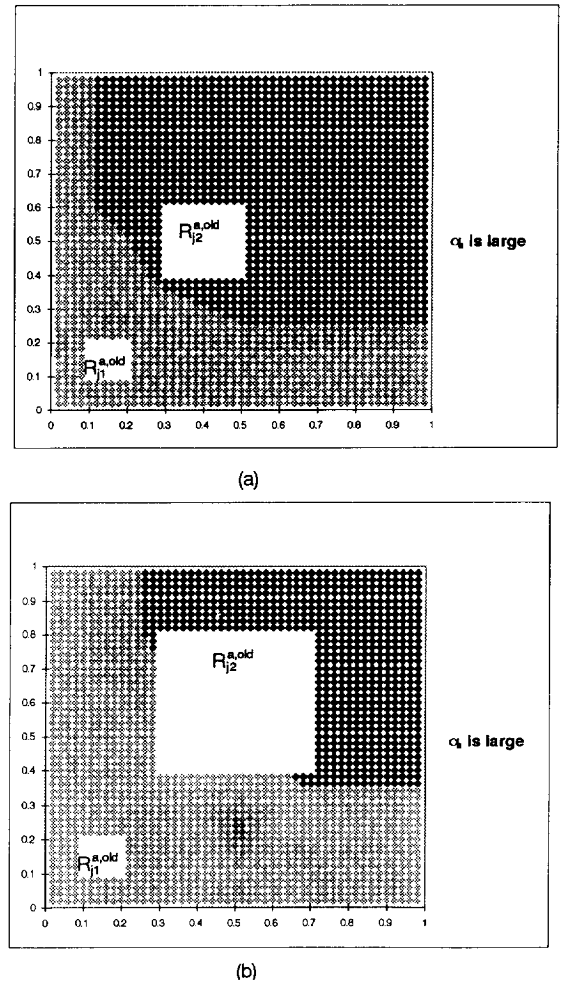


**FIGURE 14 (a,b). Illustration of the order of search rule B3b. Pattern I is outside rectangles $R_{j_1}^{a,old}$ and $R_{j_2}^{a,old}$, $|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}|$, and $\alpha_a$ is large. Pattern I chooses $R_{j_1}^{a,old}$ first if and only if $|R_{j_1}^{a,new}| < |R_{j_2}^{a,new}|$. Gray marks indicate patterns I that choose $R_{j_1}^{a,old}$ first, and black marks indicate patterns I that choose $R_{j_2}^{a,old}$ first.**

## 5. EXTENSIONS TO FUZZY ARTMAP

### 5.1. Fuzzy ARTMAP

A block diagram of the Fuzzy ARTMAP architecture is provided in Figure 16. There are many similarities between the Fuzzy ART architecture and the Fuzzy ARTMAP architecture, due to the fact that two of the three modules of Fuzzy ARTMAP are Fuzzy ART architectures. These modules are designated $ART_a$ and $ART_b$ in Figure 16. The $ART_a$ module accepts as inputs the input patterns, and the $ART_b$ module accepts as inputs the output patterns of the pattern classification task that Fuzzy ARTMAP is required to learn. All of the details mentioned about the Fuzzy ART architecture in Section 2.1 are valid for the $ART_a$ module, without any change, and for the $ART_b$ module by substituting the superscript {a} of Section 2.1 with the superscript {b} to emphasize the fact that we are referring to weights and parameter values of the $ART_b$ module. The only difference between the $ART_a$ and the $ART_b$ modules in Fuzzy ARTMAP is that for pattern classification tasks (many-to-one maps) we do not need to apply complement coding to the output patterns presented to the $ART_b$ module (see Figure 16).

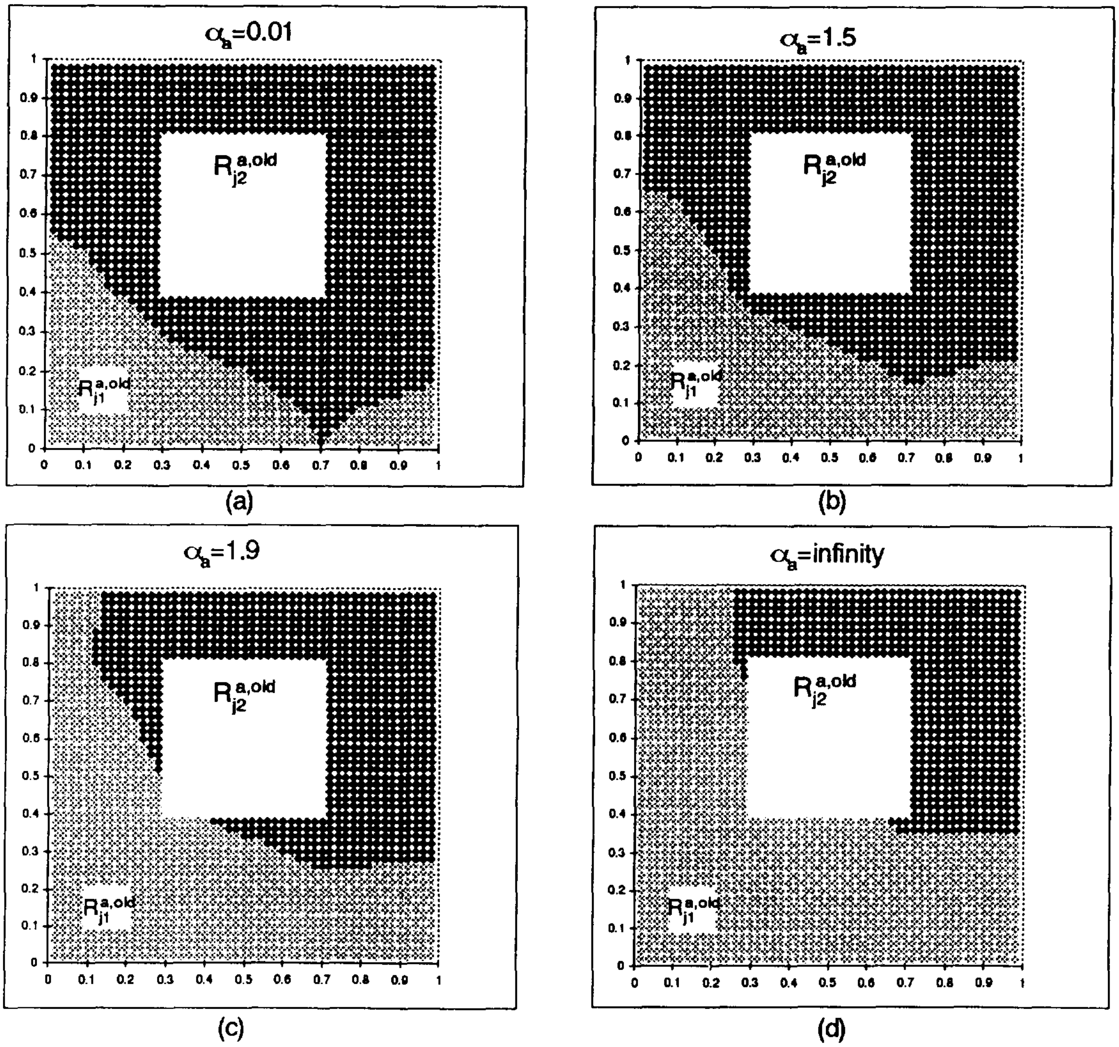


FIGURE 15 (a,b,c,d). Illustration of the order of search rule C3b. Pattern I is outside rectangles $R_{j_1}^{a,old}$ and $R_{j_2}^{a,old}$, $|R_{j_1}^{a,old}| < |R_{j_2}^{a,old}|$, and $\alpha_\alpha$ is of intermediate value. Pattern I, at times, chooses $R_{j_1}^{a,old}$ first and, at other times, it chooses $R_{j_2}^{a,old}$ first. The rectangle chosen depends on the distance of I from the rectangles and the relative sizes of the rectangles. The frequency with which rectangle $R_{j_1}^{a,old}$ is chosen first by a pattern I does not decrease as $\alpha_a$ increases from 0 to $\infty$ [compare Figure 15 (a,b,c,d)]. In the figure the gray marks indicate patterns I that choose rectangle $R_{j_1}^{a,old}$ and the black marks indicate patterns I that choose rectangle $R_{j_2}^{a,old}$.

As illustrated in Figure 16, Fuzzy ARTMAP contains a module that is designated the inter-ART module. The purpose of this module is to make sure that the appropriate mapping is established between the input patterns presented in $ART_a$ and the patterns presented in $ART_b$. There are connections (weights) between every node in the $F_2^a$ field of $ART_a$ to all the nodes in the MAP field $F_{ab}$ of the inter-ART module. The weight vector with components emanating from a node $j$ in $F_2^a$ and converging to nodes of field $F_{ab}$ is denoted $\mathbf{w}_j^{ab} = (w_{j1}^{ab}, \ldots, w_{jk}^{ab}, \ldots, w_{jN_b}^{ab})$, where $N_b$ are the number of nodes in $F_{ab}$ (the number of nodes in $F_{ab}$ is equal to the number of nodes in $F_2^b$ of $ART_b$). There are also fixed bidirectional connections between a node $k$ in $F_{ab}$ and its corresponding node $k$ in $F_2^b$.

The operation of the Fuzzy ART modules in Fuzzy ARTMAP is a little different from the operation of Fuzzy ART described in Section 2.2. For instance, resets in $ART_a$ of Fuzzy ARTMAP occur either because the category chosen in $F_2^a$ does

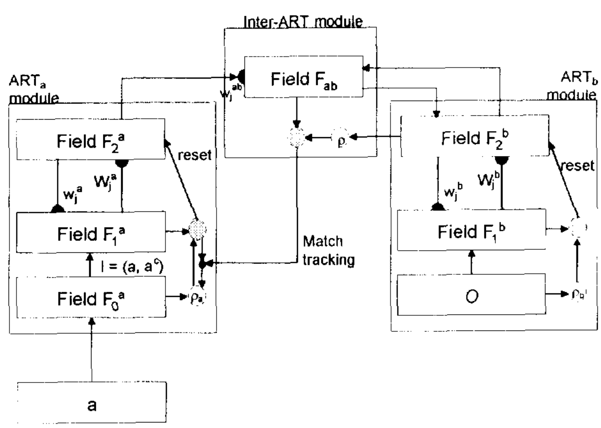


FIGURE 16. A block diagram of the Fuzzy ARTMAP architecture.

not match the input pattern presented in $F_1^a$, or because the appropriate map has not been established between an input pattern presented in $ART_a$ and its corresponding output pattern presented in $ART_b$. The latter type of reset is enforced by the inter-ART module via its connections with the orienting subsystem in $ART_a$ (see Figure 16). This reset is accomplished by forcing the $ART_a$ architecture to increase its vigilance parameter value above the level that is necessary to cause a reset of the activated node in the $F_2^a$ field of $ART_a$. Hence, in the $ART_a$ module of Fuzzy ARTMAP, we identify two vigilance parameter values, a baseline vigilance parameter value $\bar{\rho}_a$ which is the vigilance parameter of $ART_a$ prior to the presentation of an input/output pair to Fuzzy ARTMAP, and a vigilance parameter $\rho_a$ that corresponds to the vigilance parameter that is established in $ART_a$ via appropriate resets enforced by the inter-ART module of Fuzzy ARTMAP. Also, the node activated in $F_2^b$ due to a presentation of an output pattern at $F_1^b$ can either be the node receiving the maximum bottom-up input from $F_1^b$, or the node that is designated by the $F_{ab}$ field in the inter-ART module. The latter type of activation is enforced by the connections between the $F_{ab}$ field of the inter-ART module and the $F_2^b$ field of $ART_b$.

All the equations of Section 2.2 for the Fuzzy ART module are valid for the $ART_a$ and $ART_b$ module of Fuzzy ARTMAP. In particular, the bottom-up inputs to the $F_2^a$ field of the $ART_a$ module and the $F_2^b$ field of the $ART_b$ module in Fuzzy ARTMAP are given by:

$$T_j^a(\mathbf{I}) = \begin{cases} \dfrac{|\mathbf{I}|}{\alpha_a + M_u^a} & \text{if } j \text{ is an uncommitted node} \\ |\mathbf{I} \wedge \mathbf{w}_j^{a,old}|/(\alpha_a + |\mathbf{w}_j^{a,old}|) & \text{if } j \text{ is a committed node} \end{cases} \tag{73}$$

and

$$T_k^b(\mathbf{O}) = \begin{cases} \dfrac{|\mathbf{O}|}{\alpha_b + M_u^a} & \text{if } k \text{ is an uncommitted node} \\ |\mathbf{O} \wedge \mathbf{w}_k^{b,old}|/(\alpha_b + |\mathbf{w}_k^{b,old}|) & \text{if } k \text{ is a committed node} \end{cases} \tag{74}$$

where in eqn (74), $\mathbf{O}$ stands for the output pattern corresponding to the input pattern $\mathbf{I}$, while the rest of the $ART_b$ quantities are defined as they were defined for the $ART_a$ module of Fuzzy ART in Section 2.1. Similarly, the vigilance ratios for $ART_a$ and $ART_b$ are computed as follows:

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J^{a,old}|}{|\mathbf{I}|} \tag{75}$$

and

$$\frac{|\mathbf{O} \wedge \mathbf{w}_K^{b,old}|}{|\mathbf{O}|}. \tag{76}$$

The equations that describe the modifications of the weight vectors $\mathbf{w}_j^{ab}$ are explained verbally as follows. A weight vector emanating from a node in $F_2^a$ to all the nodes in $F_{ab}$ starts initially from the "all ones" vector and, after training that involves this $F_2^a$ node, all of its connections to $F_{ab}$, except one, are reduced to the value of zero.

The operating phases of Fuzzy ARTMAP are the same as the operating phases of Fuzzy ART with the only difference being that in the training phases of Fuzzy ARTMAP we present to the network input patterns along with corresponding output patterns. Also the performance of Fuzzy ARTMAP is easier to evaluate by providing the network with a test list of input/output patterns. In particular, during the performance evaluation of Fuzzy ARTMAP, only the input patterns of the test list are presented to the $ART_a$ module of Fuzzy ARTMAP. Then, Fuzzy ARTMAP makes a prediction about the corresponding output pattern and this prediction is compared with the actual corresponding output from the test list.

As we have emphasized before, since we are only focusing on pattern classification tasks, the templates formed in $ART_b$ are not very interesting (they are equal to the output patterns presented in $ART_b$). To enforce this type of clustering in $ART_b$ the vigilance parameter (i.e., $\rho_b$) in $ART_b$ is chosen equal to one. The templates formed in $ART_a$ though are a different story. The discussion in Section 2.4 about templates in Fuzzy ART is still valid for templates in the $ART_a$ module of Fuzzy ARTMAP. Furthermore, the definition of a distance in Fuzzy ART, mentioned in Section 2.5, is also valid for the $ART_a$ module of Fuzzy ARTMAP.

### 5.2. Results for Fuzzy ARTMAP

Theorems 1, 2, and 3 and Results A, B, and C of Fuzzy ART are applicable without any modification for the $ART_a$ module of Fuzzy ARTMAP. The order of search rules, established for Fuzzy ART, is also applicable for the $ART_a$ module of Fuzzy ARTMAP. As was the case with Fuzzy ART, Results A–C and the order of search rules in Section 4 explain how Fuzzy ARTMAP chooses among the committed nodes in the $ART_a$ module during *off-line training*, *on-line training*, or the *performance phase*.

Theorem 4 and Result D of Fuzzy ART are not applicable for Fuzzy ARTMAP. Theorem 4 is not valid because after the first list presentation of an off-line training phase in Fuzzy ARTMAP, the network might choose uncommitted nodes in the $ART_a$ module due to resets enforced by the inter-ART module. Resets of the inter-ART module are enforced because an input pattern from the training list is mapped to the erroneous output pattern.

## 6. REMARKS

One of our results in Section 3 referred to a Fuzzy ART architecture with $\alpha_a$ value approaching $\infty$. One might question this choice, since when $\alpha_a$ is large Fuzzy ART has the tendency to choose uncommitted nodes over existing committed nodes. Thus, it seems that Result B pertains to a Fuzzy ART architecture of questionable value, since every category formed in $F_2^a$ will correspond to a template that is equal to a pattern from the training list. This is indeed the case if we assume that the Fuzzy ART parameter $M_u^a$ is chosen equal to $2M_a$, as in the original Fuzzy ART paper (Carpenter et al., 1991b). On the other hand, if we consider a Fuzzy ART architecture with $M_u^a$ very large, so that committed nodes are chosen prior to any uncommitted node, then it is reasonable to allow $\alpha_a$ to increase to large values as well. This Fuzzy ART architecture produces a bottom-up input to a committed node $j$ in $F_2^a$, with template $\mathbf{w}_j^{a,old}$, proportional to:

$$|\mathbf{I} \wedge \mathbf{w}_j^{a,old}| \quad (77)$$

where $\mathbf{I}$ stands for the input pattern applied across the nodes of the $F_1^a$ field of Fuzzy ART. It is worth pointing out that this Fuzzy ART variant is also mentioned in Carpenter and Gjaja (1994).

To justify the large $\alpha_a$ values of the aforementioned Fuzzy ART variant ($M_u^a \to \infty$, and $\alpha_a \to \infty$), we compared its performance with the performance of the original Fuzzy ART algorithm ($M_u^a = 2M_a$, and reasonably small values for $\alpha_a$). The criterion for comparison is the average clustering performance of the algorithms. In order to do the comparison we first chose a number of benchmark databases. Then, for representative vigilance parameter values ($\rho_a$) we evaluated the average clustering performance of the Fuzzy ART variant for each of the chosen databases. For the same databases, the same representative vigilance parameter values, and a wide range of $\alpha_a$ values, we also computed the average clustering performance of the original Fuzzy ART algorithm. The results are reported in Table 1, where for ease of presentation we report the average clustering performance of the original Fuzzy ART only for selective $\alpha_a$ values; these $\alpha_a$ values gave some of the best average clustering performances of the original Fuzzy ART algorithm. In the same table, we also show the average number of $F_2^a$ nodes created by the Fuzzy ART variant and the original Fuzzy ART algorithms. Furthermore, standard deviations of the clustering performance and the number of nodes are also reported in Table 1. The results in Table 1 indicate that this Fuzzy ART variant compares favorably with the original Fuzzy ART algorithm.

In the sequel, we provide some additional details for the databases used and we explain how the average clustering performance of the Fuzzy ART variant and the Fuzzy ART algorithms was computed.

The databases used for the aforementioned comparison were: the *heart disease database*, the *diabetes database*, the *wine recognition database*, the *ionosphere database*, and the *sonar database*. These databases were selected from the collection of the databases distributed by the machine learning group at the University of California at Irvine (Murphy & Ada, 1994). The training list of each one of these databases consists of inputs and corresponding output patterns. For the training of Fuzzy ART only the input patterns of the training list were used. For the heart database, $M_a = 13$, and the training list consisted of 203 input patterns belonging to five different classes. For the diabetes database, $M_a = 8$, and the training list consisted of 576 input patterns belonging to two different classes. For the wine database, $M_a = 12$, and the training list consisted of 120 input patterns belonging to three different classes. For the ionosphere database, $M_a = 34$, and the training list consisted of 200 input patterns belonging to two different classes. For the sonar database, $M_a = 60$, and the training list consisted of 104 input patterns belonging to two different classes.

To evaluate the clustering performance of Fuzzy ART (Fuzzy ART variant or the original Fuzzy ART) we trained it with the training list of input patterns until it learned the list completely. After training was over, we assigned a label to each category formed in the $F_2^a$ field of Fuzzy ART. A category formed in $F_2^a$ of Fuzzy ART is labeled by the output pattern to which most of the input patterns that are represented by this category are mapped. Then we evaluated the performance of Fuzzy ART by presenting to it, one more time, the input patterns from the training list. For every input pattern from the training list, Fuzzy ART chooses a category in $F_2^a$. If the label of this category is the output pattern that this input pattern corresponds to in the training list, then we say that Fuzzy ART clustered this input pattern correctly. If, on the other hand, the label of this category is different from the output pattern that this input pattern corresponds to in the training list, then we say that Fuzzy ART made an erroneous clustering. It is worth mentioning that the aforementioned procedure to evaluate the performance of Fuzzy ART was motivated by Dubes and Jain (1976),

**TABLE 1**
**Average Percentage of Correct Clustering ($avg_{cl}$), Standard Deviation of Correct Clustgering ($stdv_{cl}$), Average Number of $F_2^a$ Nodes ($avg_{no}$), and Standard Deviation of $F_2^a$ Nodes ($stdv_{no}$) for Fuzzy ART with $M_u^a \to \infty$ and $\alpha_a \to \infty$, as well as Fuzzy ART with $M_u^a = 2M_a$ and Various Typical $\alpha_a$ Parameter Values**

Heart Database

| | $\rho_a = 0.3$ | | $\rho_a = 0.4$ | | | $\rho_a = 0.6$ | | $\rho_a = 0.8$ | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha_a = 6.0$ | $\alpha_a \to \infty$ | $\alpha_a = 0.01$ | $\alpha_a = 13.0$ | $\alpha_a \to \infty$ | $\alpha_a = 6.0$ | $\alpha_a \to \infty$ | $\alpha_a = 6.0$ | $\alpha_a \to \infty$ |
| Avg_cl | 57.9% | 57.7% | 61.4% | 65.1% | 62.7% | 69.3% | 70.6% | 74.6% | 77.8% |
| Stdv_cl | 1.44% | 1.64% | 1.75% | 2.04% | 2.34% | 2.13% | 1.56% | 2.08% | 1.57% |
| Avg_no | 5.50 | 5.80 | 4.00 | 15.6 | 10.3 | 23.0 | 28.6 | 55.0 | 60.4 |
| Stdv_no | 0.53 | 0.000 | 1.18 | 1.35 | 0.79 | 1.89 | 1.78 | 2.89 | 2.72 |

Diabetes Database

| | $\rho_a = 0.2$ | | $\rho_a = 0.4$ | | | $\rho_a = 0.7$ | | $\rho_a = 0.9$ | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha_a = 0.8$ | $\alpha_a \to \infty$ | $\alpha_a = 0.01$ | $\alpha_a = 7.4$ | $\alpha_a \to \infty$ | $\alpha_a = 1.0$ | $\alpha_a \to \infty$ | $\alpha_a = 2.6$ | $\alpha_a \to \infty$ |
| Avg_cl | 65.2% | 65.7% | 69.4% | 72.4% | 71.4% | 81.3% | 80.4% | 92.9% | 91.1% |
| Stdv_cl | 0.62% | 0.83% | 1.80% | 3.10% | 3.40% | 1.08% | 1.45% | 0.88% | 1.05% |
| Avg_no | 2.50 | 2.90 | 6.30 | 14.7 | 9.90 | 53.1 | 74.5 | 234.3 | 245.7 |
| Stdv_no | 0.53 | 0.32 | 0.48 | 0.95 | 0.74 | 2.02 | 3.63 | 4.22 | 2.97 |

Wine Database

| | $\rho_a = 0.5$ | | $\rho_a = 0.6$ | | | $\rho_a = 0.7$ | | $\rho_a = 0.9$ | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha_a = 5.0$ | $\alpha_a \to \infty$ | $\alpha_a = 0.01$ | $\alpha_a = 24$ | $\alpha_a \to \infty$ | $\alpha_a = 2.4$ | $\alpha_a \to \infty$ | $\alpha_a = 0.8$ | $\alpha_a \to \infty$ |
| Avg_cl | 63.2% | 68.2% | 87.8% | 90.2% | 82.8% | 92.7% | 91.0% | 98.8% | 98.2% |
| Stdv_cl | 6.98% | 6.95% | 2.74% | 4.06% | 7.73% | 2.31% | 4.42% | 0.75% | 0.86% |
| Avg_no | 3.00 | 2.00 | 5.20 | 8.80 | 5.60 | 10.0 | 11.2 | 53.6 | 53.8 |
| Stdv_no | 0.00 | 0.32 | 0.45 | 0.45 | 0.52 | 0.71 | 0.79 | 1.52 | 1.23 |

Ionosphere Database

| | $\rho_a = 0.1$ | | $\rho_a = 0.2$ | | $\rho_a = 0.4$ | | | $\rho_a = 0.7$ | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha_a = 1.2$ | $\alpha_a \to \infty$ | $\alpha_a = 12.6$ | $\alpha_a \to \infty$ | $\alpha_a = 0.01$ | $\alpha_a = 34.0$ | $\alpha_a \to \infty$ | $\alpha_a = 0.6$ | $\alpha_a \to \infty$ |
| Avg_cl | 73.2% | 89.0% | 78.7% | 93.0% | 91.9% | 87.0% | 94.6% | 94.7% | 97.2% |
| Stdv_cl | 3.30% | 2.17% | 2.99% | 1.76% | 2.69% | 2.24% | 1.07% | 2.70% | 1.53% |
| Avg_no | 6.00 | 7.70 | 17.0 | 15.0 | 26.3 | 37.1 | 31.7 | 60.6 | 63.6 |
| Stdv_no | 0.47 | 0.71 | 0.94 | 1.00 | 1.64 | 1.45 | 2.03 | 1.63 | 1.88 |

Sonar Database

| | $\rho_a = 0.4$ | | $\rho_a = 0.6$ | | | $\rho_a = 0.7$ | | $\rho_a = 0.8$ | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha_a = 0.8$ | $\alpha_a \to \infty$ | $\alpha_a = 0.01$ | $\alpha_a = 30.0$ | $\alpha_a \to \infty$ | $\alpha_a = 6.0$ | $\alpha_a \to \infty$ | $\alpha_a = 0.6$ | $\alpha_a \to \infty$ |
| Avg_cl | 54.3% | 54.9% | 93.7% | 87.0% | 86.9% | 94.4% | 91.3% | 95.4% | 92.6% |
| Stdv_cl | 2.73% | 2.24% | 2.28% | 3.30% | 5.59% | 0.99% | 3.44% | 1.91% | 1.64% |
| Avg_no | 2.00 | 2.00 | 9.80 | 9.50 | 10.2 | 18.8 | 19.5 | 33.2 | 34.4 |
| Stdv_no | 0.00 | 0.00 | 0.79 | 1.08 | 1.03 | 1.93 | 2.01 | 2.30 | 2.17 |

where a number of clustering techniques are compared with each other. The average performances, shown in Table 1, were computed by training Fuzzy ART with ten different orders of pattern presentations from the training list.

Unfortunately, the very large choices for the choice parameter value $\alpha_a$ and the parameter value $M_u^a$ are not a good combination of parameter values for Fuzzy ARTMAP. Fuzzy ARTMAP simulation results conducted with these parameter choices indicated that Fuzzy ARTMAP created too many clusters in $ART_a$ in order to solve the pattern classification tasks corresponding to the five databases, mentioned above, and as a result they are not worth mentioning here. Hence, Fuzzy ARTMAP simulations with very large $M_u^a$ and $\alpha_a$ values are not practical. Consequently, although order of search results for Fuzzy ARTMAP were derived for small, intermediate, and large $\alpha_a$ values, only the results for small and intermediate $\alpha_a$ values are of practical significance. On the other hand, the order of search results for Fuzzy ART are of practical significance for small, intermediate and large $\alpha_a$ values.

## 7. CONCLUSIONS

In this paper we investigated Fuzzy ART and Fuzzy ARTMAP from the perspective of the order according to which categories are chosen to represent the input patterns. We used the geometrical interpretation of these categories, in terms of appropriate rectangles, to try to explain these choices in a geometrical fashion. We illustrated that Fuzzy ART and Fuzzy ARTMAP exhibit three distinct behaviors, when $\alpha_a$ is small, when $\alpha_a$ is large, and when $\alpha_a$ assumes intermediate values. These behaviors were fully documented in terms of a certain order of search rules that explain how choices of categories (rectangles) are made by Fuzzy ART or Fuzzy ARTMAP.

The importance of the work conducted here lies in the fact that the geometrical interpretation of the order of choices made by Fuzzy ART and Fuzzy ARTMAP allows one to obtain a more intuitive understanding of how these algorithms operate. Furthermore, a complete understanding of the order according to which categories are chosen in Fuzzy ART or Fuzzy ARTMAP can help demonstrate useful learning properties pertaining to these two architectures. To strengthen our point, it suffices to mention that the order according to which categories are chosen in ART1 and ARTMAP (i.e., subset categories are chosen first) produced several learning properties for these architectures (Georgiopoulos et al., 1990, 1991, 1992, 1994; Huang et al., 1995; Moore 1989). One of these properties predicted a good upper bound on the number of list presentations required by ART1 and ARTMAP to learn a list of training examples repeatedly presented to either one of these two architectures.

## REFERENCES


Carpenter, G. A., & Gjaja, M. N. (1994). Fuzzy art choice functions. In *Proceedings of the World Congress on Neural Networks* (pp. I-713–722), San Diego, CA.

Carpenter, G. A., & Grossberg, S. (1987a). ART2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, **26**(23), 4919–4930.

Carpenter, G. A., & Grossberg, S. (1987b). 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. (1990). ART3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, **3**(2), 129–152.

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991a). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, **4**(5), 565–588.

Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991b). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, **4**(6), 759–771.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, **3**(5), 698–713.

Dubes, R., & Jain, A. (1976). Clustering techniques: The user's dilemma. *Pattern Recognition*, **8**, 247–260.

Georgiopoulos, M., Heileman, G. L., & Huang, J. (1990). Convergence properties of learning in ART1. *Neural Computation*, **2**(4), 502–509.

Georgiopoulos, M., Heileman, G. L., & Huang, J. (1991). Properties of learning related to pattern diversity in ART1. *Neural Networks*, **4**(6), 751–757.

Georgiopoulos, M., Heileman, G. L., & Huang, J. (1992). The N–N–N conjecture in ART1. *Neural Networks*, **5**(5), 745–753.

Georgiopoulos, M., Huang, J., & Heileman, G. L. (1994). Properties of learning in ARTMAP. *Neural Networks*, **7**, 495–506.

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

Healy, M. J., Caudell, T. P., & Smith, S. D. G. (1993). A neural architecture for pattern sequence verification through inferencing. *IEEE Transactions on Neural Networks*, **4**(1), 9–20.

Huang, J., Georgiopoulos, M., & Heileman, G. L. (1995). Fuzzy ART properties. *Neural Networks*, **8**(2), 203–213.

Moore, B. (1989). ART1 and pattern clustering. In Touretzky, D. S., Hinton, G., and Sejnowski, T. (Eds.), *Proceedings of the 1988 Connectionist Summer School* (pp. 174–185). San Mateo, CA: Morgan Kaufmann.

Murphy, P., & Ada, D. (1994). *UCI repository of machine learning databases*. Technical Report, Department of Computer Science, University of California, Irvine, CA.