Universal Approximation With Fuzzy ART and Fuzzy ARTMAP

Stephen J. Verzi
Computer Science Department, University of New Mexico
Albuquerque, NM 87131, USA
verzi@eece.unm.edu
and
Gregory L. Heileman
Department of Electrical & Computer Engineering, University of New Mexico
Albuquerque, NM 87131, USA
heileman@eece.unm.edu
and
Michael Georgiopoulos
School of Electrical Engineering and Computer Science, University of Central Florida
Orlando, FL 32816, USA
michaelg@mail.ucf.edu
and
Georgios C. Anagnostopoulos
Department of Electrical & Computer Engineering, Florida Institute of Technology
Melbourne, FL 32901, USA
georgio@fit.edu

Abstract

A measure of success for any learning algorithm is how useful it is in a variety of learning situations. Those learning algorithms that support universal function approximation can theoretically be applied to a very large and interesting class of learning problems. Many kinds of neural network architectures have already been shown to support universal approximation. In this paper, we will provide a proof to show that Fuzzy ART augmented with a single layer of perceptrons is a universal approximator. Moreover, the Fuzzy ARTMAP neural network architecture, by itself, will be shown to be a universal approximator.

Keywords: Adaptive Resonance Theory, Machine Learning, Neural Networks, Universal Function Approximation.

I. Introduction

In the late 1980's and early 1990's, important theoretical results were proved that showed certain classes of learning algorithms capable of universal function approximation. Early on it was shown that combinations of sigmoid functions could be used to support universal function approximation [1]. This result was important since a standard neural network perceptron computes a sigmoid function. Then multi-layered feedforward neural networks with either sigmoid or Gaussian kernel functions were shown to be universal approximators [2], [3], [4]. Also, radial basis function neural networks were proved to be capable of universal function approximation [5]. Very recently a hybrid ART-based neural network has been shown to be a universal approximator [6].

In this paper we will show that Fuzzy ART with only an extra layer of perceptrons can support universal approximation. More importantly, the Fuzzy ARTMAP neural network by itself can perform universal approximation. The result showing Fuzzy ART to be a universal approximator is an important fact in establishing the utility of ART-based neural network architectures as viable learning techniques. A learning algorithm which is known to be a universal approximator can be applied to a large class of interesting problems with the confidence that a solution is at least theoretically available.

Before presenting our main results, we will describe the Fuzzy ART and Fuzzy ARTMAP neural network architectures. Then we will present a proof showing how a Fuzzy ART neural network extended with a layer of perceptrons can be used to support universal function approximation. Next we will show how Fuzzy ARTMAP by itself can support this same capability. Finally, we will discuss the utility of Fuzzy ART, the curse of dimensionality and practical learning algorithms.

II. Fuzzy ART and Fuzzy ARTMAP

Fuzzy ARTMAP is a neural network architecture designed to learn a mapping between example instances and their associated labels [7]. These training examples are denoted (x, y), where $x \in [0,1]^m$ is an example data instance, and $y \in [0,\infty)^d$ is its corresponding d-dimensional label. Fuzzy ARTMAP is composed of two Fuzzy ART neural network modules connected through a MAP field, as shown in Fig. 1.

During training, the pair (x, y) is preprocessed to form the pair $((x\ x^c), (y\ y^c))$ which is then presented to the neural network. The instance x is presented to the A-side Fuzzy ART module (ART^A) and label y is presented to the B-side Fuzzy ART module (ART^B) in Fig. 1. Fuzzy ARTMAP performs supervised learning by enforcing that the A-side F_2 node which learns x will only be associated with a single B-side F_2 which learns y.

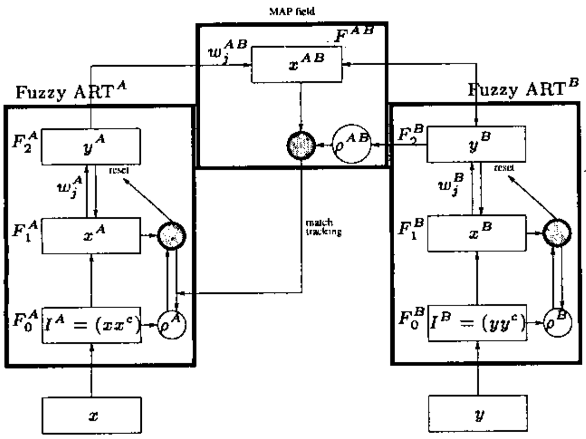

Fig. 1. The Fuzzy ARTMAP Architecture.

A. Fuzzy ART

The Fuzzy ART neural network architecture was designed to conduct unsupervised learning, or clustering, on real-valued data [8]. Clustering is the process of grouping similar data points into *cluster groups*, *clusters* or *categories*. The measure of similarity among data points for Fuzzy ART will be specified in detail below. In this section, we will describe the structure of the Fuzzy ART neural network architecture, followed by a detailed look at its clustering algorithm.

Fuzzy ART Input Data. The Fuzzy ART neural network architecture, with complement coding, assumes that the input data used to train it is normalized to fit within the unit hypercube. Thus, an input data point, x, is a m-dimensional vector of values each of which lies between zero and one, inclusive (i.e., $x_i \in [0,1], i = 1,2,\ldots,m$). The dimension, m, is constant for a particular learning problem. The complement of an input vector, x, is now well defined as $x_i^c = 1 - x_i, i = 1,2,\ldots,m$.

Fuzzy ART Structure. Fuzzy ART is structured into three layers of interacting nodes, labeled F_0, F_1 and F_2, where the output of F_0 is connected to F_1, and F_1 and F_2 are mutually interconnected, as shown in Fig. 1. At F_0, a m-length input vector from the environment is complement coded and passed on to F_1. The process of *complement coding* a pattern vector, x, produces a new vector $I = (x, x^c)$, where x^c is the complement of x defined previously.

There are $2m$ nodes in layer F_1, and $N \geq 1$ nodes in layer F_2. Each node in the F_2 layer is fully connected, by a weighted link, w_j in Fig. 1, to each node in the F_1 layer.[1] The number of nodes in the F_2 layer is allowed to grow as necessary during learning. An F_2 layer node that has learned at least one data point is called *committed*. A Fuzzy ART neural network module always has one uncommitted node in the F_2 layer available for training, along with $N-1$ committed nodes. When the uncommitted node learns its first data point, a new uncommitted node then becomes available, and N is increased by one. Each committed F_2 node and its associated weights, w_j represents a separate category of input data, also called a category *template*.

The output vector, y, from a Fuzzy ART network consists of boolean values signifying those F_2 nodes which are active. Thus

$$y_j = \begin{cases} 1, & \text{if } F_2 \text{ node j is active} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $1 \leq j \leq N$. Note that the uncommitted node, N, is available in (1). The operation of Fuzzy ART ensures that only a single F_2 node is active for a given pattern. We will see this in more detail in the algorithmic description presented next.

Fuzzy ART Algorithm. The Fuzzy ART algorithm described here is a combination of work presented by Carpenter and Moore, however, the symbols used will reflect those used throughout the rest of this paper for consistency [8], [9]. For a given input data point, the Fuzzy ART learning algorithm has three stages. First, the input is complement coded. Then the "best" matching F_2 node is found for the complement coded input data. Note that the F_2 node found might be the uncommitted node, and initially, a Fuzzy ART neural network architecture has only the single uncommitted node available for learning. Finally, the best matching F_2 node found is allowed to learn the new data point. Given a complement coded input vector, I, the similarity measure at node j of the F_2 layer, called $T_j(I)$, is computed as a weighted sum of I and the weights w_j, shown in (2). Note that these weights connect the F_1 layer nodes to node j in the F_2 layer.

The mathematical formula used by Fuzzy ART to find the best matching category template during cluster formation is

$$J = \arg \max_{0 \leq j \leq N} T_j(I), \tag{2}$$

where

$$T_j(I) = \begin{cases} \frac{|I \wedge w_j|}{\alpha + |w_j|}, & \text{if } \frac{|I \wedge w_j|}{|I|} \geq \rho \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The parameter α, called the choice parameter, is usually a small positive quantity, $\wedge$ is the element-wise vector *min* operator, and $|\cdot|$ is the L_1-norm of a vector. The best matching F_2 node from the choice competition, J, must satisfy the vigilance criterion

$$\frac{|I \wedge w_J|}{|I|} \geq \rho. \tag{4}$$

The vigilance parameter, ρ in (4), is a user-supplied input between zero and one. Note that at least one F_2 node, the

[1] Actually, there are bottom-up and top-down weights connecting the nodes in the F_1 layer to the nodes in the F_2 layer. In this paper, the top-down weights representing the cluster template are the only weights of interest, and so these weight will be referred to as w_j.

uncommitted node, will always satisfy the vigilance criterion. The maximum choice F_2 template node satisfying the vigilance criterion is allowed to learn the input vector, a condition called *resonance.* Ties between F_2 nodes with the same choice value are broken by assigning an index to all F_2 nodes, and choosing the node with the lowest index value in a tie. The index values are assigned when F_2 nodes are committed.

Initially all template weights w_j are set to one, and learning proceeds as follows

$$w_J^{(new)} = \beta(I \wedge w_J^{(old)}) + (1-\beta)w_J^{(old)}, \qquad (5)$$

where β is the learning parameter. In this paper we use $\beta = 1$, which is a special case called *fast learning.* Note that learning only occurs at the winning F_2 node, J, during resonance. An important feature of Fuzzy ART is that the F_2 layer grows as needed for a particular problem.

Fuzzy ART F_2 Node Category Template. The Fuzzy ART neural network module accepts a vector of values as input, but it also produces a vector of values as output. A committed Fuzzy ART F_2 node j has a weight vector defined as $w_j = x_1 \wedge x_2 \wedge \ldots \wedge x_n$, where F_2 node j has learned all of the input data points in $X = \{x_1, x_2, \ldots, x_n\}$. Because of complement coding, w_j defines the minimum hyperbox containing the data points in X. The vigilance criterion ensures that $|w_j| \geq \rho$.

$$|w_j| = \sum_{i=1}^{2m} w_{ji} = \sum_{i=1}^{2m} \min_{k=1}^{n} x_{ki} \geq \rho \qquad (6)$$

Thus, $w_j = (pq^c)$ where $p_k = \min_{i\in\{1,2,\ldots,n\}} x_{ik}$ and $q_k = \max_{i\in\{1,2,\ldots,n\}} x_{ik}$. The axis-parallel hyper-rectangle for w_j has a minimum point at p and a maximum point at q. The first m points from w_j are the "lower left" corner, and the second m points are the complement of the "upper right" corner of the hyperbox defined by the F_2 node j. The vigilance parameter, ρ, can be used to control the granularity of clusters covering the problem space. A larger ρ value will force Fuzzy ART to create smaller clusters, necessitating more clusters to cover a larger problem space. A smaller ρ value will allow Fuzzy ART to create larger clusters, meaning fewer clusters are needed to cover a problem space.

B. Fuzzy ARTMAP

The Fuzzy ARTMAP architecture shown in Fig. 1 consists of two Fuzzy ART modules connected by a MAP field. The ART^A module is given pattern data and the ART^B module is given label data for a given supervised learning task. The MAP field links data cluster templates (A-side) with label cluster templates (B-side). Supervised learning is performed in Fuzzy ARTMAP by ensuring that each ART^A template is linked with only one ART^B template. Thus, a many-to-one association from pattern to label templates is formed in the Fuzzy ARTMAP MAP field.

The Fuzzy ARTMAP MAP field weights, w_{jk}^{AB}, are used to control associations between A-side F_2 nodes and B-side F_2 nodes. An uncommitted A-side F_2 node, j, has the following initial weight values

$$w_{jk}^{AB} = 1, \ \forall k, \ 0 \leq k \leq N^B, \qquad (7)$$

meaning that j is not currently associated with any B-side F_2 node (there are N^B B-side F_2 nodes), and in fact it is available for future learning. An uncommitted A-side F_2 node j becomes committed with B-side F_2 node K through the following weight assignments

$$w_{jK}^{AB} = 1 \text{ and } w_{jk}^{AB} = 0, \ \forall k \neq K, \qquad (8)$$

thus A-side F_2 node, j, is exclusively and permanently linked with B-side F_2 node, K.

The Fuzzy ARTMAP architecture ensures the many-to-one mapping through the use of a match tracking lateral reset, as shown in Fig. 1. The lateral reset is used in Fuzzy ARTMAP to ensure that each training pattern resonates with an A-side F_2 node associated with a B-side F_2 node that is consistent with the pattern's label. After a bounded number of epochs, Fuzzy ARTMAP is guaranteed to reach a steady state [10]. Note that during testing it is possible for a test pattern, never seen before, to choose the uncommitted node. In this case no B-side label prediction is possible.

III. Main Results

In this chapter, we will present a proof showing the universal approximation capabilities of the Fuzzy ART neural network. Actually, these results will apply to a modified Fuzzy ART network. It will also be shown that the Fuzzy ARTMAP network can be used without further modification to perform universal approximation.

In order to show that the Fuzzy ART F_2 node is capable of universal approximation, it will be necessary to show that the Fuzzy ART neural network architecture is capable of computing any member of a sequence of functions, $S = \bigcup s_n(x)$ which are dense in $L^p(\Re^m)$. Actually, Fuzzy ART with complement coding operates in the unit hypersquare, thus the space of interest is $L^p([0,1]^m)$ where $[0,1]^m \subset \Re^m$. Saying that $s_n(x)$ is dense in $L^p([0,1]^m)$ where $1 \leq p < \infty$ is equivalent to saying that for every $f \in L^p([0,1]^m)$ and every $\epsilon > 0$, there exists $\phi \in S$ such that $\|f - \phi\|_p \leq \epsilon$. If we have $S \subset L^p([0,1]^m)$ and our modified Fuzzy ART architecture can be shown to compute any member of S, then we will have shown that this neural network is capable of universal function approximation in $L^p([0,1]^m)$.

Fuzzy ART F_2 nodes conduct data clustering similar to the internal layer nodes of a radial basis function (RBF) neural network [5]. By itself then, the Fuzzy ART module cannot be expected to perform function approximation, but by adding an output layer of perceptron nodes, this approximation can be achieved. The internal layer of an

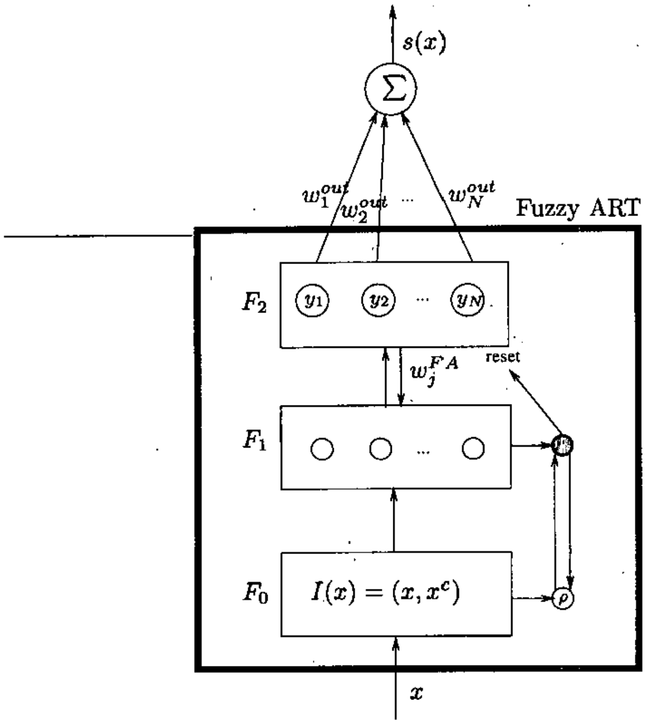

Fig. 2. Modified Fuzzy ART for Universal Approximation.

RBF neural network is connected to an output layer of nodes to perform approximation in a similar manner, see Figure 1 in [5]. In Fig. 2, the Fuzzy ART neural network module is connected to an output layer perceptron. Universal approximation for the Fuzzy ART will apply to an m-dimensional input space, but for simplicity in this paper, only a single input dimension will be considered. Also, the approximation output will, in general, fit multi-dimensional output, but for simplicity, only a single output dimension will be considered, and thus, only a single output layer perceptron is used, as shown in Fig. 2.

There are several steps necessary to prove the universal approximation capabilities of the Fuzzy ART F_2 node. Given a measurable function $f \geq 0$, the first task will be to determine a sequence of functions which will be used to approach f from below. This sequence of functions will rely upon partitioning of the domain of f into disjoint sets. Next, the Fuzzy ART F_2 node will be shown to be capable of computing the indicator function for an arbitrary member of these disjoint sets. It is this indicator function that will actually be used in the sequence of functions that we are interested in. Finally, these results will be pulled together with the construction and specification of the modified Fuzzy ART neural network architecture, shown in Fig. 2, for computing the sequence of functions for approximating f, and this sequence of functions will be shown to be dense in $L^p([0,1]^m)$.

Given a measurable function $f \in L^p([0,1]^m)$, $f \geq 0$ and $\epsilon > 0$, Consider the following sequence of functions for $n = 1, 2, 3, \cdots$

$$
\begin{aligned}
s_n(x) &= \sum_{j=1}^{2^n} C_{n,j}\chi_{D_{n,j}}(x) \qquad (9)\\
D_{n,j} &= \left[\frac{j-1}{2^n}, \frac{j}{2^n}\right),\ 1 \leq j < 2^n \\
D_{n,2^n} &= \left[\frac{2^n-1}{2^n}, 1\right],
\end{aligned}
$$

where the coefficients $C_{n,j}$ are determined using f and the diadic sets $D_{n,j} \subset [0,1]$. Diadic sets, also called diadic intervals, boxes or cubes are members of $\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3 \cup \cdots$, where Ω_n is defined as the collection of all 2^{-n} boxes with corners at P_n, and P_n is the set of all $x \in \Re^m$ whose coordinates are integer multiples of 2^{-n} [11]. Thus, $D_{n,j} \in \Omega_n$. Ω's density in $\Re^m$ and its capacity to cover (measurable) sets is exploited by the sequence of functions defined in (10) [11], [12]. Later it will be shown that $s_n(x) \leq f(x)$ for all x except for a set of measure less than ϵ. Note that $s_n(x)$ are simple functions if the coefficients take on a finite number of values, and this will be shown below. Thus $s_n(x) \in L^p([0,1]^m)$.

Next, it will be shown that the modified Fuzzy ART neural network architecture in Fig. 2 can be configured such that it computes $s_n(x)$ in (10). The first step is to show that F_2 nodes in Fig. 2 can compute the indicator functions of $D_{n,j}$ from (10) with proper network quantities including weights, indexes and vigilance values).

Lemma 1: The Fuzzy ART F_2 node can compute the indicator function for $D_{n,j} \in \Omega_n$.

Proof. Given $D_{n,j}$, the network quantities for a Fuzzy ART F_2 node will be determined so that it computes $\chi_{D_{n,j}}(x)$. A Fuzzy ART F_2 node has three quantities that need to be determined, w_j^{FA}, ρ and the node index V_j. The minimum point in $D_{n,j}$ is $\frac{j-1}{2^n}$ and the minimum point not in $D_{n,j}$ is $\frac{j-1}{2^n} + 2^{-n} = \frac{j}{2^n}$. Therefore, the weights become

$$
w_j^{FA} = \left(\frac{j-1}{2^n} \ \left(\frac{j}{2^n}\right)^c\right) \qquad (10)
$$

The function computed by the Fuzzy ART F_2 node j is

$$
\begin{aligned}
y_j(x) &= \begin{cases} 1 & \text{if } j = \arg\max_{1\leq i\leq N} T_i(I(x)) \\ 0 & \text{otherwise} \end{cases} \qquad (11)\\
I(x) &= (x\ x^c) \\
T_j(I) &= \begin{cases} \frac{\|w_j^{FA} \wedge I\|_1}{\|w_j^{FA}\|_1 + \alpha} & \text{if } \frac{\|w_j^{FA} \wedge I\|_1}{\|I\|_1} \geq \rho \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

where $I(x)$ is the complement coded value for x and α is a small positive number. Note that because of complement coding $\| I(x) \|_1 = 1$, $\forall x$ (in general this will be m, but here $m = 1$). The vigilance parameter will have the following value

$$
\rho = \frac{\| w_j^{FA} \|_1}{m} = \| w_j^{FA} \|_1 \qquad (12)
$$

Thus, $y_j(x)$ will only be 1 if $x \in \left[\frac{j-1}{2^n}, \frac{j}{2^n}\right]$. And so the Fuzzy ART F_2 node j computes the indicator function for the closed interval $\left[\frac{j-1}{2^n}, \frac{j}{2^n}\right]$ and F_2 node $j+1$ will compute the indicator function for $\left[\frac{j}{2^n}, \frac{j+1}{2^n}\right]$. Note that these two overlap at $\frac{j}{2^n}$. In Fuzzy ART ties between competing F_2 nodes are broken by assigning an index value to the competing F_2 nodes and choosing the F_2 node with the lowest index. The index values, V_j, for the modified Fuzzy ART F_2 nodes become $V_j = 2^n - j + 1$. Thus, $y_j(x) = \chi_{\left[\frac{j-1}{2^n}, \frac{j}{2^n}\right)}(x)$ for $1 \leq j < 2^n$, and $y_{2^n}(I(x)) = \chi_{\left[\frac{2^n-1}{2^n}, 1\right]}(x)$. Therefore, the Fuzzy ART F_2 nodes constructed as described above compute the disjoint intervals $D_{n,j}$ for $1 \leq j \leq 2^n$ from (10). $\Diamond$

Before proceeding with the construction of the modified Fuzzy ART network, the coefficients $C_{n,j}$ from (10) will be specified. Here is the complete specification of the sequence of functions $s_n(x)$ including the coefficients

$$\begin{aligned} s_n(x) &= \sum_{j=1}^{2^n} C_{n,j}\chi_{D_{n,j}}(x) \qquad (13)\\ C_{n,j} &= \frac{A_{n,j}-1}{2^n} \\ A_{n,j} &= \min_{i \in B_{n,j}} (i) \\ B_{n,j} &= \left\{ i : \mu(E_{n,i} \cap D_{n,j}) \geq \frac{\epsilon}{2^n} \right\}, \\ & \quad 1 \leq i \leq n2^n + 1 \\ D_{n,j} &= \left[\frac{j-1}{2^n}, \frac{j}{2^n}\right), \ 1 \leq j < 2^n \\ D_{n,2^n} &= \left[\frac{2^n-1}{2^n}, 1\right], \\ E_{n,i} &= f^{-1}\left(\left[\frac{i-1}{2^n}, \frac{i}{2^n}\right)\right), \ 1 \leq i \leq n2^n \\ E_{n,n2^n+1} &= f^{-1}([n, \infty)). \end{aligned}$$

Note that $E_{n,i}$ are the pre-images of the Lebesgue intervals of f [11]. The index values in $B_{n,j}$ refer to those $E_{n,i}$ which intersect with $D_{n,j}$ with measure greater than or equal to $\frac{\epsilon}{2^n}$. Therefore, $C_{n,j} \leq f(x)$ for all $x \in D_{n,j}$ except for a set of measure $< \frac{\epsilon}{2^n}$. Note that there are 2^n diadic sets $D_{n,j}$. And so, $s_n(x) \leq f(x)$ for all x except for a set of measure ϵ.

Now the final result can be shown.

Theorem 1: The modified Fuzzy ART neural network, shown in Fig. 2, can be used to universally approximate any measurable function in $L^p([0,1])$.

Proof. Given $1 \leq p < \infty$ and $f \in L^p([0,1])$, $f \geq 0$, a series of functions, s_n, computable by the modified Fuzzy ART neural network in Fig. 2, will be determined such that these functions approximate f in the limit, and it will be shown that s_n is dense in $L^p([0,1])$. The Fuzzy ART neural network shown in Fig. 2 with parameters determined in (10) and (12) computes the following function for the $N = 2^n$ F_2 nodes

$$\begin{aligned} s_n(x) &= \sum_{j=1}^{N} w_{n,j}^{out} \cdot y_{n,j}(x) \qquad (14)\\ y_{n,j}(x) &= \begin{cases} 1 & \text{if } j = \arg\min_{k \in J_n} V_k \\ 0 & \text{otherwise} \end{cases} \\ J_n &= \left\{ i : i = \arg\max_{1 \leq j \leq N} T_{n,j}(I(x)) \right\} \\ T_{n,j}(I) &= \begin{cases} \frac{\|w_{n,j}^{FA} \wedge I\|_1}{\|w_{n,j}^{FA}\|_1 + \alpha} & \text{if } \frac{\|w_{n,j}^{FA} \wedge I\|_1}{\|I\|_1} \geq \rho \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Given the results from Lemma 1, these equations can be reduced to

$$\begin{aligned} s_n(x) &= \sum_{j=1}^{N} w_{n,j}^{out} \cdot \chi_{D_{n,j}}(x) \qquad (15)\\ D_{n,j} &= \left[\frac{j-1}{2^n}, \frac{j}{2^n}\right), \ 1 \leq j < 2^n \\ D_{n,2^n} &= \left[\frac{2^n-1}{2^n}, 1\right] \end{aligned}$$

The final step is to determine the values for $w_{n,j}^{out}$, which can be set as

$$w_{n,j}^{out} = C_{n,j} \qquad (16)$$

where $C_{n,j}$ is defined in (14).

Since $|f(x) - s_n(x)|^p \leq f^p$, Lebesgue's dominated convergence theorem implies $||f(x) - s_n(x)||^p \to 0$ as $n \to \infty$ [11]. Our modified Fuzzy ART neural network computes the function in (14). Since $0 \leq s_n \leq f$, then $s_n \in L^p([0,1]^m)$. Thus, f is in the L^p-closure of s_n. $\Diamond$

We have implemented a very simple, proof of concept, version of the modified Fuzzy ART architecture in MATLAB™, as shown in Fig 3. The modified Fuzzy ART architecture is capable of representing any measurable function in $L^p([0,1])$ with an arbitrarily large number of F_2 nodes. Note that for each n in the sequence

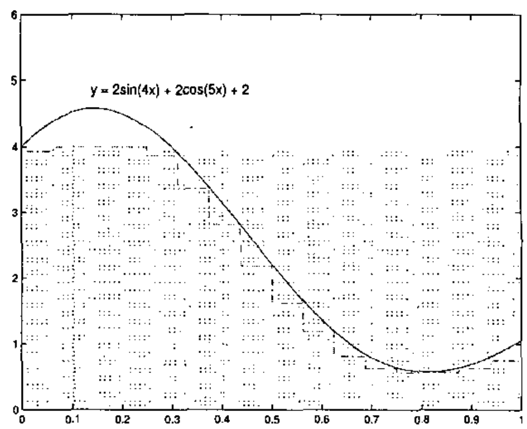

Fig. 3. Modified Fuzzy ART with $n = 4$.

defined above, a separate Fuzzy ART network is needed since there is only a single vigilance parameter ρ. Fuzzy ART F_2 nodes, in a single network, can represent any $D_{n,j} \in \Omega_n$.

A. Universal Approximation With Fuzzy ARTMAP

It is a very simple extension to use Fuzzy ARTMAP, by itself, instead of the modified Fuzzy ART neural network to perform universal approximation. With Fuzzy ARTMAP, we need to construct the A-side Fuzzy ART module, the B-side Fuzzy ART module and the MAP field. The A-side Fuzzy ART module will be constructed precisely as described in the previous section, however, instead of using the extra layer of perceptron nodes with their association weights, we use the Fuzzy ARTMAP MAP field and the B-side category templates. There will be one B-side template for each of the Lebesgue intervals, $E_{n,i}$ in (14). The value of the B-side template weight will be exactly the same as the coefficients used in the modified Fuzzy ART network, $C_{n,j}$ in (14). Note that there will be no complement coding in the B-side Fuzzy ART module. Next, the MAP field will be used to compute the minimum intersection between the diadic cube that A-side node J computes and the pre-images of all B-side nodes $1 \leq k \leq N^B$, $A_{n,j}$ in (14). Therefore, $w_{jk}^{AB} = 1$ if $A_{n,j} = k$, otherwise $w_{jk}^{AB} = 0$. Because each $A_{n,j}$ is unique, then the constructed MAP field will conform to Fuzzy ARTMAP learning in that each A-side F_2 node is associated with only a single B-side F_2 node. Thus, given input x, the value output by this Fuzzy ARTMAP network will be $C_{n,j}$ where $x \in D_{n,j}$.

IV. Concluding Remarks

An example of an ART-based neural network architecture that can perform such universal approximation described above is BARTMAP-SRM [13]. BARTMAP-SRM operates using the diadic hyperboxes described previously. The network is initiated using the unit hyperbox, which is subsequently split in half across each dimension creating 2^m new squares for an m-dimensional input learning space. Note that BARTMAP-SRM as well as the network constructions described in our main results above both suffer from the curse of dimensionality [14]. This means that for an m-dimensional input space, an exponentially large number of internal layer F_2 nodes may be required to reach a final solution. BARTMAP-SRM was not designed as a practical solution for high dimension input problems but rather as a theoretical construct for demonstrating the universal approximation capabilities. Another modification to Fuzzy ART, called Boosted ART has been shown to represent the same data space with exponentially fewer F_2 nodes, with respect to the input dimension m [15]. It is hoped that Boosted ART can be used in high dimension input spaces to help address the curse of dimensionality.

In this paper we have shown that Fuzzy ART augmented with a single layer of perceptron nodes can support universal function approximation. Furthermore, we have demonstrated that Fuzzy ARTMAP, by itself is a universal approximator. These results establish both of these neural network architectures as viable learning techniques on a large class of learning problems. These results continue to suffer from the curse of dimensionality, as do other universal approximation results. Our future research continues to expand upon these results in designing practical ART-based neural network architectures for conducting learning.

Acknowledgment

The authors from the University of Central Florida acknowledge the partial support from NSF through a CRCD grant number 0203446 entitled "Machine Learning Advances for Engineering Education".

References

[1] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematical Journal of Control, Signals, and Systemics*, vol. 2, pp. 303–314, 1989.

[2] K. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks*, vol. 2, pp. 183–192, 1989.

[3] K. M. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.

[4] E. J. Hartman, J. D. Keeler, and J. M. Kowalski, "Layered neural networks with gaussian hidden units as universal approsimators," *Neural Compuation*, vol. 2, pp. 210–215, 1990.

[5] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Computation*, vol. 3, pp. 245–257, 1991.

[6] L. Martí, A. Policriti, and L. García, "AppART: An ART hybrid stable learning neural network for universal function approximation," in *Hybrid Information Systems*, A. Abraham and M. Koeppen, Eds., Heidelberg, Jul 2002, pp. 92–120, Physica Verlag.

[7] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 698–713, 1992.

[8] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, no. 5, pp. 759–771, 1991.

[9] B. Moore, "ART 1 and pattern clustering," in *Proceedings of the 1988 Connectionist Summer School*. 1988, pp. 174–185, Morgan Kaufmann.

[10] M. Georgiopoulos, J. Huang, and G. L. Heileman, "Properties of learning in ARTMAP," *Neural Networks*, vol. 7, no. 3, pp. 495–506, 1994.

[11] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, second edition, 1974.

[12] E. DiBenedetto, *Real Analysis*, Birkhäuser, Boston, 2001.

[13] S. J. Verzi, G. L. Heileman, M. Georgiopoulos, and G. C. Anagnostopoulos, "Off-line structural risk minimization and BARTMAP-S," in *Proceedings of the International Joint Conference on Neural Networks*, 2002.

[14] H. White, *Artificial Neural Networks: Approximation and Learning Theory*, Blackwell, Cambridge MA, 1992.

[15] S. J. Verzi, *Boosted ART and Boosted ARTMAP: Extensions of Fuzzy ART and Fuzzy ARTMAP*, Ph.D. thesis, University of New Mexico, Albuquerque, New Mexico, 2002.