COMPARATIVE PERFORMANCE MEASURES OF FUZZY ARTMAP, LEARNED VECTOR QUANTIZATION, AND BACK PROPAGATION FOR HANDWRITTEN CHARACTER RECOGNITION

Gail Carpenter[†], Stephen Grossberg[‡], and Kunihiko Iizuka [§]

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

ABSTRACT

This article compares the performance of Fuzzy ARTMAP with that of Learned Vector quantization and Back Propagation on a handwritten character recognition task. Training with Fuzzy ARTMAP to a fixed criterion used many fewer epochs. Voting with Fuzzy ARTMAP yielded the highest recognition rates.

1. Introduction

Fuzzy ARTMAP (Carpenter, Grossberg, Markuzon, Reynolds, and Rosen, 1991) has been proposed as a neural network architecture for supervised learning of recognition categories and multidimensional maps in response to arbitrary sequences of analog or binary input vectors. The first ARTMAP system (Carpenter, Grossberg, and Reynolds, 1991) was used to classify binary vectors. The generalization to analog input vectors is accomplished by replacing the ART 1 modules in ARTMAP by Fuzzy ART modules (Carpenter, Grossberg, and Rosen, 1991). In this work, we use handwritten digit recognition as a test problem for comparing the performance of Fuzzy ARTMAP with other two commonly used neural networks, Learned Vector Quantization (LVQ) and Back Propagation (BP). Fuzzy ARTMAP was superior to LVQ and BP in terms of the speed of learning and prediction accuracy.

2. Summary of the Fuzzy ARTMAP Algorithm

Fuzzy ARTMAP consists of a pair of Fuzzy ART modules (ART_a and ART_b) linked together via an inter-ART associative memory F^{ab} that is called a *map field*. During supervised learning, the Fuzzy ARTMAP receives a stream $(\mathbf{a}_p, \mathbf{b}_p)$ of input patterns, where $\mathbf{b}_p$ is the correct prediction given $\mathbf{a}_p$. The ART_a and ART_b modules classify $\mathbf{a}_p$ and $\mathbf{b}_p$ into
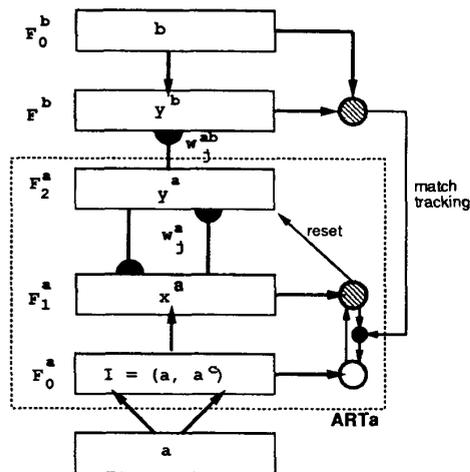
Figure 1: Simplified Fuzzy ARTMAP architecture for classification problem. The map field and ART$_b$ of the original Fuzzy ARTMAP are reduced to a field F^b where each cell corresponds to a pattern class.

categories and the map field makes associations from ART$_a$ categories to ART$_b$ categories. If a_p predicts an incorrect b_p, then the mismatch between actual and predicted b_p causes a memory search within ART$_a$ by a mechanism called *match tracking*. Match tracking realizes a Minimax Learning Rule which conjointly minimizes predictive error and maximizes generalization under fast, incremental learning conditions in response to arbitrarily ordered input pairs (a_p, b_p). Match tracking increases the ART$_a$ vigilance parameter ρ_a by the minimum amount needed to trigger a memory search. Vigilance relaxes to its baseline vigilance $\bar{\rho}$ between learning trials. For a pattern classification problem where each b_p itself represents the class to which a_p belongs, the ART$_b$ and the map field can be replaced by a single field F^b (Figure 1).

3. Handwritten Digit Data Base and Preprocessing

The data base consists of 600 examples of digits which are handwritten by 30 persons. Each person wrote each digit twice using a tablet and a stylus pen. The position of the stylus pen on the tablet was sampled at a speed of 5 points per second. Then the sampled points were thinned out to compress the data without substantially degrading character appearance, and added to the data base. Stroke features were extracted from each of the 4 x 4 rectangular receptive fields $A_1, \ldots, A_{16}$ (Figure 2). In each receptive field A_i, we calculated the projections (H_i, V_i, R_i, L_i) of the part of stroke in A_i onto four representative orientation (horizontal, vertical, and two diagonal) lines as well as the number E_i of stroke end points which fall into A_i. They were combined into an 80-dimensional feature vector $(H_1, V_1, R_1, L_1, E_1, \ldots, H_{16}, V_{16}, R_{16}, L_{16}, E_{16})$. Then each vector generated from 600 examples of the data was normalized to form a vector **a** whose components range from 0.0 to 1.0.

4. Simulation Results and Comparison with Other Methods

In each simulation, we used 400 examples randomly chosen from the data base for learn-
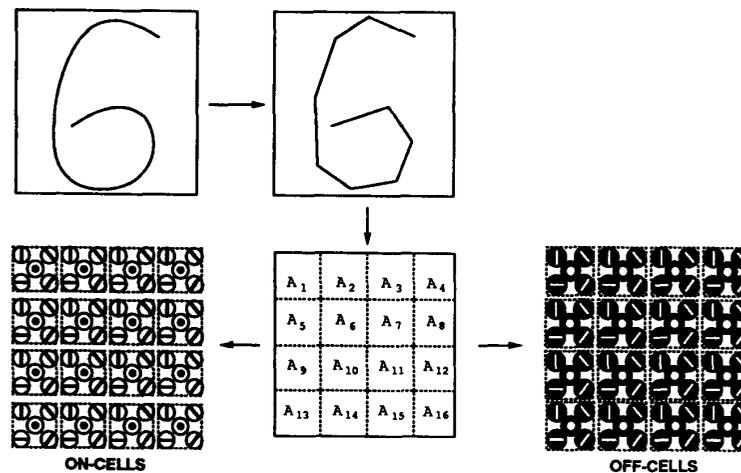
Figure 2: Preprocessing for Fuzzy ARTMAP with complement coding. Each output cell represents an orientation sensitive or end point sensitive on-cell or off-cell.

No. Epochs	1	5	10	20	50	100	200
10 hidden units	83.5	91.0	93.0	94.5	94.5	96.0	96.0
20 hidden units	65.5	90.5	93.5	92.5	93.5	93.5	93.0

Table 1: Correct test set prediction rates (%) of BP network with 10 or 20 hidden units for various number of learning epochs.

No.Reference Vectors	10	20	30	40	50	60	70	80
	89.5	94.5	93.0	94.5	95.5	96.5	95.0	94.5

Table 2: Correct test set prediction rates (%) of LVQ network for various number of reference vectors.

ing and the remaining 200 examples to test the predictive generalization of the trained networks. We compared Fuzzy ARTMAP performance with that of the LVQ and BP neural network classifiers on the same data base using the same preprocessing to generate input vectors **a**. The BP network had 80 input units, 10 or 20 hidden units, and 10 output units with an adaptable bias for each hidden and output unit. The best test set prediction rate of the BP network was 96.0% with 10 hidden units. It decreased to 94.0% as the number of hidden unit increases to 20 (Table 1). Table 2 summarizes test set and training set prediction rates on 8 LVQ simulations along with the numbers of reference vectors. For the LVQ learning, we started with reference vectors generated by k-mean methods and evolved them through 500 x (the number of reference vectors) randomly chosen training pattern presentations.

Table 3 shows a series of Fuzzy ARTMAP simulation results corresponding to various

	α	ρ	Correct Test Set Prediction Rate(%)	No.ARTa Categories	No.Epochs
	0.1	0.0	95.2(94.5-96.0)	21.6(19-26)	2.0(2-2)
	1.0	0.0	95.9(94.5-96.5)	23.2(20-29)	2.8(2-3)
(a)	5.0	0.0	95.9(92.0-98.5)	28.8(26-31)	4.2(3-5)
	10.0	0.0	95.8(94.5-98.0)	42.2(36-50)	6.0(5-7)
	20.0	0.0	94.2(91.5-96.0)	66.6(56-81)	8.4(7-12)
	100.0	0.0	90.4(86.0-93.5)	172.2(116-208)	14.6(12-19)
	0.1	0.5	87.1(84.0-90.5)	28.6(24-34)	2.0(2-2)
	1.0	0.5	87.4(84.5-91.0)	29.6(26-36)	2.4(2-3)
(b)	5.0	0.5	95.4(92.5-98.5)	31.2(25-36)	3.6(3-5)
	10.0	0.5	96.3(95.9-98.0)	43.2(36-50)	6.0(5-8)
	20.0	0.5	94.2(91.5-96.0)	66.6(56-81)	8.4(7-12)
	100.0	0.5	90.4(86.0-93.5)	172.2(116-208)	14.6(12-19)
	0.1	0.6	83.5(79.5-88.5)	42.0(40-45)	2.0(2-2)
	1.0	0.6	84.1(80.5-89.5)	42.6(40-48)	2.2(2-3)
(c)	5.0	0.6	87.8(82.0-92.5)	44.4(41-51)	2.6(2-3)
	10.0	0.6	91.9(89.9-94.0)	45.0(40-48)	4.0(3-5)
	20.0	0.6	95.4(92.5-97.0)	70.0(65-77)	8.2(7-9)
	100.0	0.6	90.4(86.0-93.5)	172.2(116-208)	14.6(12-19)

Table 3: **Fuzzy ARTMAP** simulations with $\bar{\rho}$ equal to (a) 0.0, (b) 0.5, and (c) 0.6 for diverse α. Each result shows the average over 5 independent simulation results as well as their performance range denoted in parentheses.

combination of choice parameter α and baseline vigilance $\bar{\rho}$. The choice parameter α appears in the functions $T_j = | \, \mathbf{I} \wedge \mathbf{w}_j^a \, | \, (\alpha + | \, \mathbf{w}_j^a \, |)^{-1}$ that are defined for each category j in F_2^a (Figure 1). The category j whose T_j is maximal is selected to represent input **a**. The statistics are averages from 5 simulations using different randomly chosen test and training sets. The training patterns were presented in a different order for the number of epochs shown in Table 3 until 100% training set performance was reached.

The peak performances are 95.9% at $\alpha = 1.0$ or $\alpha = 5.0$ for $\bar{\rho} = 0.0$, 96.3% at $\alpha = 10.0$ for $\bar{\rho} = 0.5$ and 95.4% at $\alpha = 20.0$ for $\bar{\rho} = 0.6$. Figure 3 shows the prediction rate as a function of α. Although these rates are almost identical, the required number of committed nodes and training epochs increases with α. Baseline vigilance $\bar{\rho} = 0$ uses fewer ART$_a$ categories and generates stable performance for a relatively wide range of α.

Table 4 shows how test set performance changes depending on the number of training epochs of Fuzzy ARTMAP with $\alpha = 1.0$, 5.0 and 10.0. After one epoch, for any α, the correct prediction rate on the test set reaches above 99% of its best performance.

Figure 4 shows examples of category templates and a part of coded patterns which Fuzzy ART$_a$ module with $\bar{\rho} = 0.0$ and $\alpha = 1.0$ has learned through one epoch presentation of training patterns.

The prediction rate of Fuzzy ARTMAP can be improved by using the voting strategy (Carpenter, Grossberg, Markuzon, Reynolds, and Rosen, 1991). Here several ARTMAPs are trained on different random orderings of the same inputs, and the category favored by the
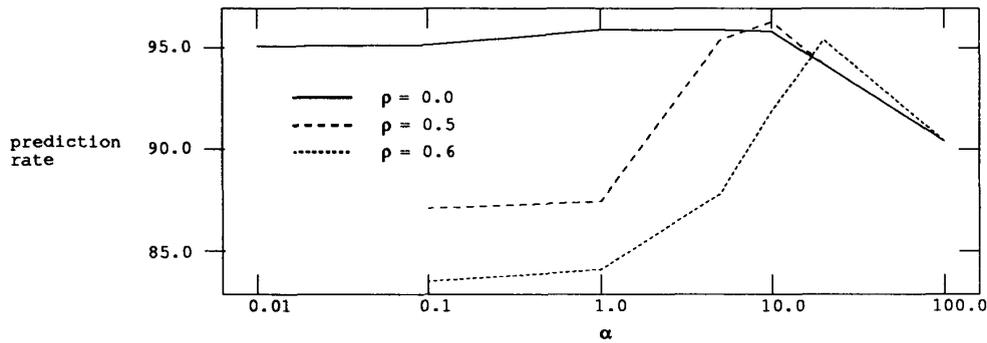
Figure 3: The correct prediction rate vs. choice parameter α for three different values of vigilance parameter $\bar{p}$.
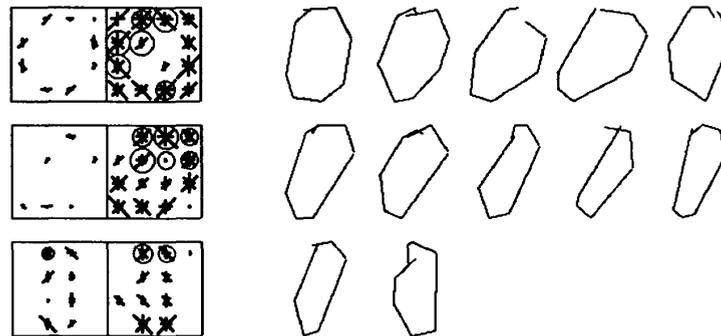
Figure 4: Fuzzy ARTMAP category templates and coded patterns. The left box represents on-cell adaptive weights and the right box represents the complementary off-cell weights, corresponding to three Fuzzy ART categories. Illustrative exemplars of the digit "0" from each category are displayed.

majority is selected. Table 5 shows how voting improves performance of Fuzzy ARTMAP. Each result of 5-voter Fuzzy ARTMAP performance is an average of five independent 5-simulation votings. In each case, 5-simulation voting eliminated 24-36% of the test set error, yielding the best prediction accuracy of the three models. Moreover, the fluctuation in prediction rates are reduced to 1-2% from 2.1-6.5% for 1-simulation systems. This makes the system highly reliable.

In summary:

1. The best correct prediction rates for BP, LVQ, and a single Fuzzy ARTMAP are almost equivalent. ARTMAP voting generates the best prediction accuracy of all the models.

2. In terms of the number of training epochs needed to achieve a certain predictive rate, Fuzzy ARTMAP is far superior to BP and LVQ. Even one epoch of Fuzzy ARTMAP training yields close to maximal performance.

3. In terms of the total number of pathways needed for the best performance, the BP network is superior to Fuzzy ARTMAP and to LVQ. However, all compression in ARTMAP occurs between layers F_1^a and F_2^a in Figure 1. For each F_2^a category, only one weight to F^b

	α	ρ	Correct Test Set Prediction Rate(%)	No.ARTa Categories	No.Epochs
(a)	1.0	0.0	95.9(94.5-96.5)	22.8(19-28)	1
	1.0	0.0	95.9(94.5-96.5)	23.2(20-29)	2
(b)	5.0	0.0	95.2(92.0-96.5)	26.0(23-29)	1
	5.0	0.0	96.1(93.0-97.5)	28.0(26-30)	2
	5.0	0.0	96.0(92.5-97.5)	28.6(26-31)	3
(c)	10.0	0.0	94.0(92.0-96.5)	32.6(31-35)	1
	10.0	0.0	95.6(94.0-98.0)	36.2(32-40)	2
	10.0	0.0	95.3(94.0-96.5)	38.6(35-42)	3
	10.0	0.0	95.6(94.0-97.5)	40.4(35-46)	4
	10.0	0.0	95.7(94.0-98.0)	41.6(36-48)	5

Table 4: The change of correct prediction rate of Fuzzy ARTMAP with α equal to (a) 1.0, (b) 5.0, and (c) 10.0 along with the increase of number of training set presentation epochs. Each result shows the average over 5 independent simulation results as well as their performance range denoted in parentheses.

α	ρ	Correct Test Set Prediction Rate(%)	No.ARTa Categories	No.Epochs
5.0	0.0	97.4(96.5-98.5)	30.0(25-36)	4.6(3-7)
5.0	0.5	96.7(95.0-97.5)	32.2(26-42)	3.5(2-5)
10.0	0.5	97.2(97.0-98.0)	41.8(32-50)	5.8(3-8)

Table 5: The performances of 5-voter Fuzzy ARTMAP with ($\alpha = 5$ and $\bar{\rho} = 0$), ($\alpha = 5$ and $\bar{\rho} = 0.5$), and ($\alpha = 10$ and $\bar{\rho} = 0.5$). Each result shows the average of 5 independent 5-simulation votings as well as their performance range denoted in parentheses.

is non-zero and many weights from F_1^a to F_2^a equal zero. In contrast, BP uses all three levels to build its map. For example, in the best BP simulations, 92% of weights from level 2 to level 3 are at least 5% of the maximal weight value in absolute size.

References

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. (1991). Fuzzy ARTMAP:A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transaction on Neural Networks*, **3**, (in press).

Carpenter, G.A., Grossberg, S. and Reynolds, J.H. (1991). ARTMAP:Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, **4**, 565-588. Reprinted in: Carpenter, G.A. and Grossberg, S. (Eds.) (1991) **Pattern Recognition by Self-Organizing Neural Networks**, Cambridge, MA: MIT Press.

Carpenter, G.A., Grossberg, S. and Rosen, D.B. (1991). Fuzzy ART:Fast stable learning and categorization of analog patterns. *Neural Networks*, **4**, 759-771.