Fuzzy ARTVar: An Improved Fuzzy ARTMAP Algorithm

I. Dagher*, M. Georgiopoulos*, G. L. Heileman**, G. Bebis***

* Department of Electrical and Computer Engineering
University of Central Florida,Orlando, FL 32816
Phone: (407)823-5338 Fax: (407)823-5835
mng@ece.engr.ucf.edu

** Department of Electrical and Computer Engineering
University of New Mexico, Albuquerque, NM 87131
Phone: (505)277-4011 Fax: (505)277-1439
heileman@eece.unm.edu

*** Department of Computer Science
University of Nevada-Reno, Reno, NV 89557
Phone: (702)784-6463 Fax: (702)784-1877
bebis@cs.unr.edu

Abstract

In this paper we introduce a variation of the performance phase of Fuzzy ARTMAP which is called *Fuzzy ARTVar*. Experimental results have shown that Fuzzy ARTVar exhibits superior generalization performance, compared to Fuzzy ARTMAP, for a variety of machine learning databases. Furthermore, experimental results have also demonstrated that Fuzzy ARTVar compares favorably with other existing variations of Fuzzy ARTMAP, such as ARTEMAP (power rule), ARTEMAPQ (Q-max rule), and Gaussian ARTMAP. What is worth noting is that the performance of Fuzzy ARTVar is independent of the tuning of network parameters, in contrast with the ARTEMAP, ARTEMAPQ, and Gaussian ARTMAP algorithms, whose performance depends on the choice of certain network parameters.

1 Introduction

Pattern classification is a key element to many engineering solutions. Sonar, radar, seismic, and diagnostic applications all require the ability to accurately classify data. Control, tracking and prediction systems will often use classifiers to determine input-output relationships. Because of this wide range of applicability, pattern classification has been studied a great deal. Simpson in his Fuzzy Min-Max paper (see [1]) identified a number of desirable properties that a pattern classifier should possess. These properties are listed below: **Property 1:** *On-Line Adaptation*, **Property 2:** *Non-Linear Separability*, **Property 3:** *Short Training Time*, **Property 4:** *Soft and Hard Decisions*, **Property 5:** *Verification and Validation*, **Property 6:** *Independence from Tuning Parameters*, **Property 7:** *Nonparametric Classification*, and **Property 8:** *Overlapping Classes.*

A neural network classifier that satisfies most of these properties is Fuzzy ARTMAP ([2]). Fuzzy ARTMAP is a member of the class of neural network architectures referred to as *ART-architectures* developed by Carpenter, Grossberg, and their colleaugues at Boston University. The ART-architectures are based on the *ART theory* introduced by Grossberg in ([3]). In this paper, we focus on a variation of the performance phase of the Fuzzy ARTMAP algorithm that provides, in many instances, an improved performance. We refer to this variation of the Fuzzy ARTMAP algorithm as *Fuzzy ARTVar.* It is also worth mentioning that recently, modifications of the performance phase of the Fuzzy ARTMAP algorithm have appeared in the literature (e.g., ARTEMAP (power rule in [4]), ARTEMAPQ (Q-max rule in [4]), and ARTMAP-IC [5]) that improved the performance of Fuzzy ARTMAP. One of the important differences between Fuzzy ARTVar and ARTEMAP, ARTEMAPQ and ARTMAP-IC is that the performance of the latter three algorithms depends on network parameters (p for ARTEMAP, Q for ARTEMAPQ, and τ and Q for ARTMAP-IC), while Fuzzy ARTVar does not. Hence, Fuzzy ARTVar satisfies one of the important properties of a classifier system, that is independence from tuning parameters, while ARTEMAP, ARTEMAPQ, and ARTMAP-IC do not. Note that in the simulations reported in this paper the choice parameter (β_a) and the baseline vigilance parameter ($\bar{\rho}_a$) of Fuzzy ARTVar, ARTEMAP and ARTEMAPQ are chosen equal to zero.

The organization of the paper is as follows: In Section 2, we discuss briefy the Fuzzy ARTMAP architec-

0-7803-4859-1/98 $10.00©1998 IEEE

ture, its training phase, and its performance phase. In Section 3, we introduce the modification of the performance phase of the Fuzzy ARTMAP algorithm, that leads us to the algorithm that we called *Fuzzy ARTVar*. In Section 4, we experimentally demonstrate the superiority of Fuzzy ARTVar versus Fuzzy ARTMAP for a number of databases. In the same section, we provide performance comparisons between Fuzzy ARTVar, ARTEMAP, ARTEMAPQ, and Gaussian ARTMAP for the same set of databases. Gaussian ARTMAP is an ART-based algorithm ([6]) whose training phase and performance phase differ from the corresponding phases in Fuzzy ARTMAP, but Gaussian ARTMAP's operations in the training phase resemble the operations of Fuzzy ARTVar in its performance phase. We see in Section 4, that Fuzzy ARTVar compares very favorably with ARTEMAP, ARTEMAPQ, and Gaussian ARTMAP, despite the fact that each one of these algorithms depends on the choice of a parameter, whose optimum value is database dependent. In Section 5, we provide a review of the paper and some conclusive remarks.

2 The Fuzzy ARTMAP Neural Network

A detailed description of the Fuzzy ARTMAP neural network can be found in [2]. For completeness, in the following, we present only the necessary details.

The Fuzzy ARTMAP neural network consists of two Fuzzy ART modules, designated as ART_a and ART_b, as well as an inter-ART module as shown in Figure 1. Inputs ($\mathbf{I}$'s) are presented at the ART_a module, while their corresponding outputs ($\mathbf{O}$'s) are presented at the ART_b module. The inter-ART module includes a MAP field whose purpose is to determine whether the correct mapping has been established from inputs to outputs.

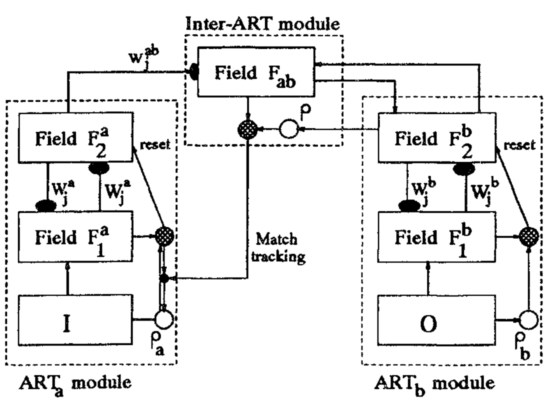

Figure 1: The Fuzzy ARTMAP Neural Network.

Fuzzy ARTMAP can operate in two distinct phases: the *training phase* and the *performance phase*. In this paper we focus on classification tasks, where many inputs are mapped to a single, distinct output. It turns out that for classification tasks , the operations performed at the ART_b and inter-ART modules can be ignored, and the algorithm can be described by referring only to the top-down weights and the parameters of the ART_a module.

The *training phase* of Fuzzy ARTMAP works as follows: Given a list of training input/output pairs, such as $\{\mathbf{I}^1, \mathbf{O}^1\}, \ldots \{\mathbf{I}^r, \mathbf{O}^r\}, \ldots \{\mathbf{I}^{PT}, \mathbf{O}^{PT}\}$, we want to train Fuzzy ARTMAP to map every input pattern of the training list to its corresponding output pattern. In order to achieve the aforementioned goal, we present the training list repeatedley to the Fuzzy ARTMAP architecture. That is present $\mathbf{I}^1$ to ART_a and $\mathbf{O}^1$ to ART_b, then $\mathbf{I}^2$ to ART_a and $\mathbf{O}^2$ to ART_b, and finally $\mathbf{I}^{PT}$ to ART_a and $\mathbf{O}^{PT}$ to ART_b; this corresponds to *one list presentation.* We present the training list as many times as it is necessary for Fuzzy ARTMAP to correctly classify all the input patterns. The classification task is considered accomplished (i.e., the learning is complete) when the weights do not change during a list presentation. The aforementioned training scenario is called *off-line learning.* Note that an input pattern $\mathbf{I}$ presented to module ART_a of Fuzzy ARTMAP is complementary encoded. That is $\mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$, where $\mathbf{a}$ is a pattern of arbitrary dimensionality with components in the interval $[0, 1]$, and $\mathbf{a}^c = \mathbf{1} - \mathbf{a}$, where $\mathbf{1}$ is the all ones vector of the same dimensionality as $\mathbf{a}$ and $\mathbf{a}^c$.

The *performance phase* of Fuzzy ARTMAP works as follows: Given a list of test input patterns, such as $\tilde{\mathbf{I}}^1, \ldots, \tilde{\mathbf{I}}^2, \ldots, \mathbf{I}^{\tilde{P}S}$, we want to find the Fuzzy ARTMAP output produced when each one of the aforementioned test patterns is presented at its F_1^a field. In order to achieve the aforementioned goal, we present the test list once to the trained Fuzzy ARTMAP architecture.

Note that an input pattern $\mathbf{I}$ presented to module ART_a of Fuzzy ARTMAP is complementary encoded. That is $\mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$, where $\mathbf{a}$ is a pattern of arbitrary dimensionality with components in the interval $[0, 1]$, and $\mathbf{a}^c = \mathbf{1} - \mathbf{a}$, where $\mathbf{1}$ is the all ones vector of the same dimensionality as $\mathbf{a}$ and $\mathbf{a}^c$.

3 The Fuzzy ARTVar Algorithm

The training phase of the Fuzzy ARTVar algorithm is identical with the training phase of the Fuzzy ARTMAP algorithm. After the training is over in Fuzzy ARTVar, we go through another phase that we call *pre-performance* phase. The purpose of the pre-performance phase is to compute, for every committed node in F_2^a, the sample mean vector and the sample standard deviation vector of all the input training patterns that chose this node. The sample mean vector of node j in F_2^a and the sample standard deviation vector of the node j are denoted by $\mathbf{m}_j^a = (m_{j1}^a, m_{j2}^a, \ldots, m_{jM_a}^a)$ and $\mathbf{s}_j^a = (s_{j1}^a, s_{j2}^a, \ldots, s_{jM_a}^a)$, respectively. These vectors are then used in the performance phase of Fuzzy

ARTVar to produce the outputs of the test patterns.

The main difference between Fuzzy ARTVar and Fuzzy ARTMAP is that in Fuzzy ARTMAP every node (category) j in F_2^a is represented by a weight $\mathbf{w}_j^a$, or the two endpoints $\mathbf{u}_j^a$ and $\mathbf{v}_j^a$ of its corresponding hyperectangle (for more details see [7]). On the other hand, in Fuzzy ARTVar every node j is represented by its mean vector $\mathbf{m}_j^a$, and its standard deviation vector $\mathbf{s}_j^a$. It is therefore, very reasonable to refer to these values as weight values during the performance phase of Fuzzy ARTVar.

Performance Phase of Fuzzy ARTVar

1. Initialize the values of the committed weight vectors in F_2^a (i.e, the $\mathbf{m}_j^a$'s and $\mathbf{s}_j^a$'s for $0 \leq j \leq N_a - 1$) to the values that they had at the end of the pre-performance phase. A node in F_2^a is committed if it has coded at least one training input pattern during the Fuzzy ARTVar training phase.

   Also, associate every committed node in F_2^a of the trained Fuzzy ARTVar with the output pattern that it was mapped to at the end of the Fuzzy ARTVar training phase.

   Initialize the index r to the value of one.

2. Choose the r-th input pattern $\tilde{\mathbf{I}}^r = (\tilde{\mathbf{a}}^r, (\tilde{\mathbf{a}}^r)^c)$ from the test list.

3. Calculate the *Mahalanobis distance* of $\tilde{\mathbf{a}}^r$ from each $\mathbf{m}_j^a$ in F_2^a, according to the following equation. When calculating the Mahalanobis distance consider only the committed nodes in F_2^a (i.e., nodes with index j, such that $1 \leq j \leq N_a - 1$).

$$d_M(\tilde{\mathbf{a}}^r, \mathbf{m}_j^a) = (\tilde{\mathbf{a}}^r - \mathbf{m}_j^a)^T (\mathbf{\Sigma}_j^a)^{-1} (\tilde{\mathbf{a}}^r - \mathbf{m}_j^a) \qquad (1)$$

   In the above equation $\mathbf{\Sigma}_j^a$ represents the covariance matrix of the members of node j in F_2^a, with diagonal elements equal to the aforementioned variances s_{ji}^a's (calculated in the pre-performance phase), and off-diagonal elements equal to zero.

4. Choose the node in F_2^a that produces the minimum Mahalanobis distance. Assume that this node has index j_{max}. That is,

$$d_M(\tilde{\mathbf{a}}^r, \mathbf{m}_{j_{max}}^a) = \min_{1 \leq j \leq N_a - 1} d_M(\tilde{\mathbf{a}}^r, \mathbf{m}_j^a) \qquad (2)$$

   Designate the output of the presented input pattern equal to the output pattern that node j_{max} was associated to at the end of the Fuzzy ARTVar training phase.

5. If this is the last input/output pair in the test list the performance phase is considered complete. Otherwise, go to Step 2, to present the next in line input pair, by increasing the value of the index r by one.

4 Experimental Results – Comparisons

In order to demonstrate the superior performance of Fuzzy ARTVar compared to Fuzzy ARTMAP we chose to conduct experiments on a number of databases extracted from the UCI repository database ([8]). The databases chosen from the repository were: Iris, Wine, Sonar, Diabetes, Breast, Balance and Bupa. It is worth noting that the data in each of the above datasets were split into a training set (2/3 of the data) and a test set (1/3 of the data). A description of each one of these databases is provided in [8].

One of the measures of performance that we used in comparing Fuzzy ARTVar and Fuzzy ARTMAP is the *generalization performance* of these networks. The generalization performance of a network is defined to be the percentage of patterns in the test set that are correctly classified by a trained network. Since the performance of Fuzzy ARTMAP and Fuzzy ARTVar depends on the order of pattern presentation in the training set, ten different random orders of pattern presentation will be investigated, and performance measures such as the average generalization performance, the minimum generalization performance, the maximum generalization performance, and the standard deviation of the generalization performance will be produced.

Also, another measure of performance for comparing neural networks is the size of the networks created. In order to compare the sizes of the networks that Fuzzy ARTVar and another algorithm create (e.g., Fuzzy ARTMAP) we compare the average compression ratio of the other algorithm versus the average compression ratio of Fuzzy ARTVar. The average compression ratio for Fuzzy ARTVar (other algorithm) is defined to be the ratio of the average number of nodes created in F_2^a versus the number of patterns used in the training of Fuzzy ARTVar (other algorithm).

In some of the databases that we experimented with we ended up, for some of the categories in F_2^a, with a standard deviation of zero across a certain dimension. Since in Fuzzy ARTVar, the criterion of choosing a node in F_2^a during the presentation of a test pattern is the minimization of (1), nodes with zero variances across some dimension will never be chosen (because then the corresponding covariance matrix inverses will be infinite). To alleviate this problem if some node variances were found to be zero, we substituted these zero variances with the minimum of the positive variance corresponding to this node. The resulting algorithm, we named *Fuzzy ARTVArc*. From now, when we refer in the main text to Fuzzy ARTVar we will imply either Fuzzy ARTVar or Fuzzy ARTVarc.

4.1 Comparisons of Fuzzy ARTVar and Fuzzy ARTMAP

Table 1 shows a comparison of the generalization performances of Fuzzy ARTMAP and Fuzzy ARTVar for the seven databases described above. Observing the results depicted in Table 1, we can draw the following conclusions.

1. In **all** of the databases the average generalization performance of Fuzzy ARTVar is **better** than the average generalization performance of Fuzzy ARTMAP. In particular, for the Bupa, Balance, Diabetes, Sonar the average generalization performance improvement is 2.9%, 4.04%, 6.04%, and 6.96%, respectively. What makes the above statement even stronger is that the standard deviation of the generalization performances is **better** for Fuzzy ARTVar for **five out of the seven** databases (Iris, Wine, Sonar Diabetes, and Breast), while it is worse for the other two (Balance and Bupa).

2. In **all** of the databases the minimum generalization performance of Fuzzy ARTVar is **equal or better** (equal only in one database) than the minimum generalization performance of Fuzzy ARTMAP. In particular, for the Bupa, Diabetes, and Sonar databases the minimum generalization performance of Fuzzy ARTVar is better than the minimum generalization performance of Fuzzy ARTMAP by 4.38%, 7.45% and 11.59%, respectively.

   In **all** of the databases the maximum generalization performance of Fuzzy ARTVar is **better** than the maximum generalization performance of Fuzzy ARTMAP. In particular, for the Bupa, Diabetes, and Balance databases the maximum generalization performance of Fuzzy ARTVar is better than the maximum generalization performance of Fuzzy ARTMAP by 4.38%, 4.71% and 6.73%, respectively.

The generalization performance in Fuzzy ARTMAP can be improved by increasing the value of the baseline vigilance parameter ($\bar{\rho}_a$). Table 2, illustrates the generalization performance comparisons between the optimum Fuzzy ARTMAP (Fuzzy ARTMAP with the optimum $\bar{\rho}_a$ value) and Fuzzy ARTVar. In Table 2, the entries B, W, and E deliver the information that, with respect to the associated performance measure (e.g., average generalization), Fuzzy ARTVar performs Better, Worse, or Equally well as the best Fuzzy ARTMAP does. Note that the compression ratio in Table 2 is always in favor of Fuzzy ARTVar. Numerical comparisons between the Fuzzy ARTVar and the optimum Fuzzy ARTMAP are not included due to lack of space.

4.2 Comparisons of Fuzzy ARTVar, with ARTEMAP, ARTEMAPQ, and Gaussian ARTMAP

As we have emphasized in the Introduction, the performance of ARTEMAP, ARTEMAPQ, and Gaussian ARTMAP depend on the values of the network parameters p, Q and γ, respectively. Furthermore, the optimum parameter value for these algorithms is database dependent. Finally, sometimes, the performance of these algorithms for the same database varies widely for different parameter values. Hence, a fair comparison between Fuzzy ARTVar, and the aforementioned algorithms is not possible since the Fuzzy ARTVar performance does not depend on any network parameter value. Nevertheless, we performed a comparison of Fuzzy ARTVar and the optimum ARTEMAP (ARTEMAP with optimum p value), ARTEMAPQ (ARTEMAPQ with optimum Q value), and Gaussian ARTMAP (Gaussian ARTMAP with optimum γ value). Our results illustrate that Fuzzy ARTVar compares favorably with the optimum ARTEMAP, optimum ARTEMAPQ, and optimum Gaussian ARTMAP. For illustration purposes we only show (in Table 3) the comparison between Fuzzy ARTVar and optimum ARTEMAPQ. In Table 3, the entries B, W, and E deliver the information that, with respect to the associated performance measure (e.g., average generalization), Fuzzy ARTVar performs Better, Worse, or Equally well as optimum ARTEMAPQ does. Numerical comparisons between the Fuzzy ARTVar and the optimum ARTEMAPQ are not included due to lack of space.

5 Review – Conclusions

We introduced a variation of the performance phase of Fuzzy ARTMAP, that we called *Fuzzy ARTVar*. We demonstrated that for a number of classification problems the performance of Fuzzy ARTVar is superior to the performance of Fuzzy ARTMAP (see Table 1). We have also implemented other variations of the performance phase of Fuzzy ARTMAP that have appeared in the literature (such as ARTEMAP, ARTEMAPQ), as well as the Gaussian ARTMAP algorithm. In comparing the existing algorithms ARTEMAP, ARTEMAPQ, and Gaussian ARTMAP with Fuzzy ARTVar we observed that Fuzzy ARTVar compares favorably with each one of these algorithms despite the fact that the performance of these algorithms was optimized with respect to an appropriate network parameter. It is also worth noting that we discovered databases (e.g., the ionosphere database) for which Fuzzy ARTVar performed worse than Fuzzy ARTMAP. Knowing though there are very few instances that a new classification algorithm will outperform existing classification algorithms for every possible classification problem, we believe that Fuzzy ARTVar is a good algorithm to consider in conjuction with or instead of algorithms such as Fuzzy ARTMAP, ARTEMAP, ARTEMAPQ, or Gaussian ARTMAP.

References

[1] P. K. Simpson, "Fuzzy min-max neural networks – Part 1: Classification," *IEEE Transactions on Neural Networks*, Vol. 3, No. 5, pp. 776-786, September 1992.

[2] 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, September 1992.

[3] S. Grossberg, "Adaptive pattern recognition and universal recoding II: Feedback, expectation, olfaction, and illusions, *Biological Cybernetics*, Vol. 23, pp. 187-202, 1976.

[4] G. A. Carpenter, and W. D. Ross, "ART-EMAP: A neural network architecture for object recognition by evidence accumulation," *IEEE Transactions on Neural Networks*, Vol. 6, No. 4, pp. 805-818, July 1995.

[5] G. A. Carpenter, and N. Markuzon, "ARTMAP-IC and medical diagnosis: Instance counting and inconsistent cases," *Technical Report CAS/CNS-96-017*, Boston University, May 1996.

[6] J. R. Williamson, "Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps," *Neural Networks*, Vol. 9, No. 5, pp. 881-897, 1996.

[7] M. Georgiopoulos, H. Fernlund, G. Bebis, and G. L. Heileman, "Order of search in Fuzzy ART and Fuzzy ARTMAP: Effect of the choice parameter," *Neural Networks*, Vol. 9, No. 9, pp. 1541-1559, 1996.

[8] P. Murphy, and D. Ada, UCI Repository of Machine Learning databases, *Technical Report*, Department of Computer Science, University of California, Irvine, CA, 1994.

Table 1: Comparison of Fuzzy ARTMAP and Fuzzy ARTVar generalization performances

Database	network	minimum	maximum	average	std. dev.
Iris	Fuzzy ARTMAP	89.58	95.83	**95.00**	1.91
	Fuzzy ARTVar	93.75	97.92	**96.67**	1.38
Wine	Fuzzy ARTMAP	91.38	98.28	**95.69**	2.70
	Fuzzy ARTVar	94.83	100	**97.93**	1.69
Sonar	Fuzzy ARTMAP	63.77	78.26	**70.58**	4.15
	Fuzzy ARTVar	75.36	79.71	**77.54**	1.75
Diabetes	Fuzzy ARTMAP	61.57	70.98	**66.63**	2.57
	Fuzzy ARTVar	67.06	75.69	**72.51**	2.15
	Fuzzy ARTVarc	69.02	75.69	**72.67**	1.65
Breast	Fuzzy ARTMAP	93.10	96.12	**94.35**	0.95
	Fuzzy ARTVar	90.09	96.55	**94.61**	2.23
	Fuzzy ARTVarc	94.40	96.55	**95.39**	0.75
Balance	Fuzzy ARTMAP	71.63	78.85	**75.91**	2.42
	Fuzzy ARTVar	71.63	85.58	**79.95**	4.75
Bupa	Fuzzy ARTMAP	47.37	63.16	**56.84**	4.22
	Fuzzy ARTVar	51.75	67.54	**59.74**	5.51

Table 2: Comparison of the optimum Fuzzy ARTMAP and Fuzzy ARTVar in terms of the generalization performance and compression ratio

Database	minimum	maximum	average	std. dev.	Compression Ratio
Iris	E	E	**B**	**B**	**B**
Wine	**B**	**B**	**B**	**B**	E
Sonar	W	W	W	**B**	**B**
Diabetes	**B**	**B**	**B**	**B**	**B**
Breast	W	W	W	W	**B**
Balance	W	**B**	**B**	W	**B**
Bupa	W	W	W	W	**B**

Table 3: Comparison of the generalization performances of the optimum ARTEMAPQ and Fuzzy ARTVar

Database	minimum	maximum	average	std. dev.
Iris	**B**	**B**	**B**	**B**
Wine	**B**	**B**	**B**	**B**
Sonar	**B**	**B**	**B**	**B**
Diabetes	W	W	W	**B**
Breast	**B**	**B**	**B**	W
Balance	W	**B**	**B**	W
Bupa	W	E	W	W