On the Design of an Ellipsoid ARTMAP Classifier within the Fuzzy Adaptive System ART Framework

Ross Peralta
Electrical & Computer Eng.[1]
rperalta@fit.edu

Georgios C. Anagnostopoulos
Electrical & Computer Eng.[1]
georgio@fit.edu

Eduardo Gomez-Sanchez
Comm. & Telematics[2]
edugom@tel.uva.es

Samuel Richie
Electrical Eng.[3]
richie@mail.ucf.edu

[1]Florida Institute of Technology, Melbourne, FL 32901
[2]University of Valladolid, Valladolid, Spain 47011
[3]University of Central Florida, Orlando, FL 32816

Abstract – **In this paper we present the design of Fuzzy Adaptive System Ellipsoid ARTMAP (FASEAM), a novel neural architecture based on Ellipsoid ARTMAP (EAM) that is equipped with concepts utilized in the Fuzzy Adaptive System ART (FASART) architecture. More specifically, we derive a new category choice function appropriate for EAM categories that is non-constant in a category's representation region. Additionally, we augment the EAM category description with a centroid vector, whose learning rate is inversely proportional to the number of training patterns accessing the category. Finally, we demonstrate the merits of our design choices by comparing FASART, EAM and FASEAM in terms of generalization performance and final structural complexity on a set of classification problems.**

I. Introduction

Adaptive resonance theory (ART) based neural networks constitute a large family of neural architectures that have been used in a plethora of applications ranging from data clustering, classification and function approximation tasks. They are all based on the ART paradigm first introduced in [1] and feature a variety of highly desirable properties, like the ability of incremental (online) learning, network response transparency and fast training phase. A characteristic of these networks is that they summarize the input data into clusters via the use of prototypes called categories, whose geometrical representation may vary (depending on the particular architecture) from being hyper-rectangles, hyper-spheres or hyper-ellipsoids embedded in the input space.

A member of the ART-based family is the Fuzzy Adaptive System ART (FASART) architecture, which was first presented in [2] as an enhancement to the standard Fuzzy ARTMAP (FAM) network [3]. FASART networks have also been successfully used for function approximation, data clustering, as well as classification tasks; see for example [4] and [5]. Both FAM and FASART employ categories, whose geometric representations are hyper-rectangles. However, FASART extends FAM by equipping categories with an additional centroid element and by introducing a new, parameterized category choice function (CCF). In FAM, the CCF value is constant within a category's representation region (see [6] for related definitions), while in FASART it monotonically decreases from 1 (at the centroid) to 0 beyond the boundaries of the category's representation region. Furthermore, while FAM's CCF depends on the category's size and the distance of the pattern from the category's representation region, in FASART the CCF depends on the distance of the pattern from the centroid and, implicitly, on the size of the category in a component-wise fashion. In this manner, FASART categories are appropriately defined as fuzzy sets and the CCF's value with respect to a pattern can be interpreted as its normalized, fuzzy membership in that fuzzy set. This permits the dual interpretation of FASART as a neural model as well as a formal fuzzy logic inference system, which is not the case for FAM according to [7].

Yet another ART-based architecture is Ellipsoid ARTMAP (EAM) [8]. The network shares almost all structural and behavioral features, as well as properties of learning with FAM. While FAM and FASART categories are represented as hyper-rectangles, EAM categories are of hyper-ellipsoid shape, which may be more suitable for certain learning problems. Like in the case of FAM, in EAM the CCF is of constant value within a category's representation region. In certain classification problem domains this CCF constancy may lead to unsatisfactory classification performance. More specifically, it is a known fact that patterns located inside the representation regions of two or more categories will access the category of the smallest size. This effect may potentially lead to poor approximation of the decision boundaries and could be avoided by using a CCF that is not constant within the representation region.

This paper focuses on the design of a variant of EAM, which we named FASEAM classifier. The relationship of FASEAM to EAM is the same as the one of FASART to FAM. We equip EAM categories with a centroid vector that is adjusted according to patterns accessing the categories. Furthermore, we derive a new CCF that is reminiscent (with

0-7803-9048-2/05/$20.00 ©2005 IEEE

respect to some properties) of the one used in FASART. Finally, we replace EAM's match tracking mechanism with an alternative secondary search procedure, since we discovered via experimentation that FASART's match tracking does not preserve the principle of incremental, instantaneous learning. In order to assess our design of FASEAM we performed experiments on four artificial databases, where we compared FASEAM to EAM in terms of structural complexity and generalization performance.

The rest of the paper is organized as follows. In Section II we highlight some of the main characteristics of FASART and EAM networks in terms of category descriptions. Section III talks about FASEAM's design with accompanying justifications. Section IV describes the data sets that were used in order to compare the original EAM classifier to FASEAM, our experimental settings and the results we obtained. Finally, Section V provides a brief summary of our contributions and observations.

II. FASART AND EAM CATEGORIES

A FASART category j is described by the min- and max-vectors $\mathbf{u}_j$ and $\mathbf{v}_j$ respectively as well as by their centroid vector $\mathbf{c}_j$, as depicted in Figure 1. The collection of the aforementioned vectors is called category description of j or template elements of category j. The region RR_j defined by the min- and max-vectors is called the representation region of category j (also shown in Figure 1) and represents all input patterns that are considered summarized/learned by the category. In order for a training pattern to be learned, categories compete in terms of their category choice function (CCF) values; the category featuring the highest CCF value is the winner of the competition, in which case we say that the training pattern accesses the category. Upon access (and under certain special circumstances that are not mentioned here) the category may learn the training pattern.

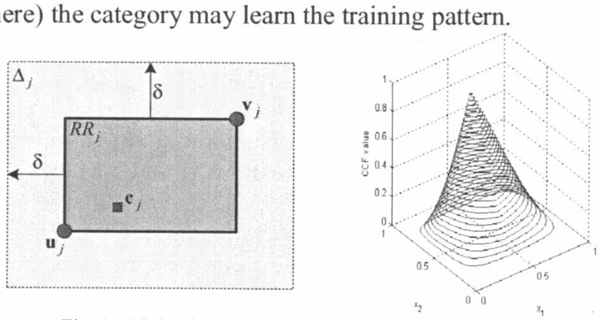

Fig. 1. Right: Geometrical depiction of a FASART category for a 2-dimensional input space. Left: Plot of the CCF values of a typical FASART category, again, for a 2-dimensional input space.

In FASART the CCF value is highest at the centroid (equal to 1) and decreases monotonically with increasing distance from the centroid $\mathbf{c}_j$ towards 0 at the boundary of the category's delta region Δ_j (also depicted in Figure 1). Note that outside Δ_j the CCF remains 0 and that the region's size is controlled by a network parameter δ (in [2] the equivalent of $\gamma=1/\delta$ is used instead).

On the other hand, EAM categories are characterized by a center $\mathbf{m}_j$, a unit-length direction vector $\mathbf{d}_j$ and a Mahalanobis radius R_j, as depicted in Figure 2. The representation region RR_j of an EAM category j is of hyper-ellipsoidal form; the hyper-ellipsoid's eccentricity is determined by a network parameter μ called axes ratio. A value of $\mu=1$ results in hyper-spherical representation regions.

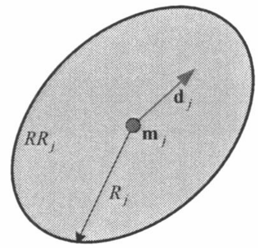

Fig. 2. Geometrical depiction of an EAM category for a 2-dimensional input space.

While in FASART distances are measured primarily using the L_1 vector norm, in EAM they are measured using a weighted Euclidian (Mahalanobis) vector norm (see Eq. 1), whose weight (shape) matrix $\mathbf{C}_j$ depends on a category's direction vector:

$$\|\mathbf{x}-\mathbf{y}\|_{\mathbf{C}_j} = \sqrt{(\mathbf{x}-\mathbf{y})^T \mathbf{C}_j (\mathbf{x}-\mathbf{y})} \tag{1}$$

$$\mathbf{C}_j = \frac{1}{\mu^2}\left[I-(1-\mu^2)\mathbf{d}_j\mathbf{d}_j^T\right] \tag{2}$$

Both network families accomplish their learning task via creation of new categories and by expanding representation regions of already existing categories. We refer the interested reader to references [2], [9] and [8] for more detailed descriptions of FASART and EAM respectively.

An advantage of FASART over FAM is that the decision boundaries created by FASART are additionally influenced by the specific location of centroid vectors, while in FAM they are influenced only by the relative shape and position of representation regions. As an example, Figure 3 compares the decision boundaries between two FAM and two FASART categories predicting different class labels.

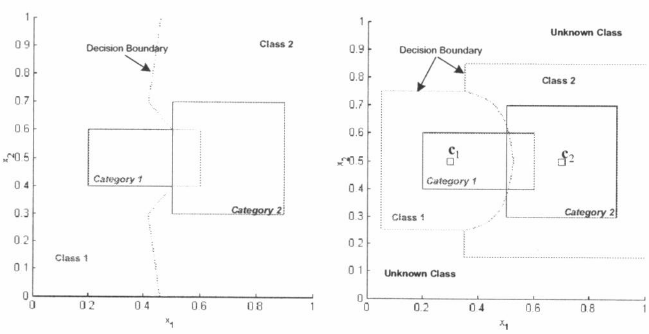

Fig. 3. Decision boundaries for two competing FAM (on the left) and two competing FASART categories (on the right).

In FAM, when two categories compete for a pattern that is inside both representation regions, the winning category is the one of the smallest size, regardless of where most of the already encoded patterns are located within the corresponding representation regions (Figure 3, right). In such a case, FASART categories, via their centroid, provide a more reasonable approach in forming smoother decision boundaries, as depicted on the right of Figure 3. As in FAM, EAM exhibits similar characteristics in decision boundary formation. This last fact has been our main motivation to derive FASEAM as an EAM variant with FASART characteristics.

III. DESIGN OF THE FASEAM CLASSIFER

Our goal is to design a variant of EAM, which we will refer to as FASEAM that resembles the original FASART architecture. As a first step, we augment the standard EAM category description with a centroid vector $\mathbf{c}_j$ and allow it to be updated by training patterns that access the category. However, instead of the standard instar learning rule for the centroid update in FASART, we update the centroid as in Gaussian ARTMAP [10]:

$$\begin{aligned} \mathbf{c}_j^{(k+1)} &= (1-\beta_c{}^{(k)})\mathbf{c}_j^{(k)} + \beta_c{}^{(k)}\mathbf{x}^{(k+1)} \\ \beta_c{}^{(k)} &\doteq \frac{1}{k+1}, \quad \mathbf{c}_j^{(1)} = \mathbf{x}^{(1)} \end{aligned} \tag{3}$$

where $\mathbf{c}_j^{(k)}$ is the updated centroid vector after the presentation of the k^{th} training pattern accessing category j. In other words, in FASEAM centroids are updated with a variable learning rate, unlike FASART centroids that are updated via a constant learning rate. The update rule in Eq. 3 forces the centroid to be the sample average of the training patterns that have accessed the category rather than a moving sample average with emphasis on the last pattern accessing the category. We chose this particular update rule to primarily enhance the stability of FASEAM during learning, since $\beta_c{}^{(k)} \le 0.5$ $\forall k \ge 2$. Let us note here that in order for the centroid to remain within the representation region of a FASEAM category after an update via Eq. 3, the learning rate β used for updating the category's center and Mahalanobis radius must be larger than 0.5.

Secondly, we have to derive a new CCF $T(j|\mathbf{x})$ that has similar properties to the one utilized in FASART while being compatible to the geometry of EAM categories. More specifically, these properties are

(i) The CCF takes values in [0,1], that is, $T(j\,|\,\mathbf{x}) \in [0,1]$

(ii) The CCF is maximum at the centroid vector and equal to 1, viz. $T(j\,|\,\mathbf{x}) = 1 \;\Leftrightarrow\; \mathbf{x} = \mathbf{c}_j$.

(iii) The CCF value is zero for any pattern outside or on the boundary of the category's delta region, i.e. $T(j\,|\,\mathbf{x}) = 0 \quad \forall \mathbf{x} \notin \operatorname{int}(\Delta_j)$

(iv) The CCF value monotonically decreases inside the delta region with increasing distance of a training pattern from the category's centroid, viz. $\dfrac{dT(j\,|\,\mathbf{x})}{d\|\mathbf{x}-\mathbf{c}_j\|} < 0 \quad \forall \mathbf{x} \in \Delta_j$, where $\|\;\|$ denotes any vector norm. Here, it is also implicitly assumed that $T(j|\mathbf{x})$ is continuous.

Although conditions (i)-(iv) reflect the behavior of FASART's CCF, they are not sufficient to uniquely determine the CCF to be used for FASEAM. Nevertheless, we start with the assumption that the constant T (CCF value) locus for a FASEAM category j should be a hyper-ellipsoid of a given center $\mathbf{z}$ and a Mahalanobis radius r, both of which depend on the specific value of T. In particular, we assume that the points $\mathbf{x}$ of the input space, that would feature a specific value $T \in [0,1]$, satisfy the equation

$$\|\mathbf{z}(T) - \mathbf{x}\|_{C_j} = r(T) \tag{4}$$

where

$$\begin{aligned} \mathbf{z}(T) &\doteq T\mathbf{c}_j + (1-T)\mathbf{m}_j \\ r(T) &\doteq (1-T)R'_j \end{aligned} \tag{5}$$

In the above equations $\mathbf{m}_j$, $\mathbf{c}_j$, R_j and $\mathbf{C}_j$ are the center vector, centroid vector, Mahalanobis radius and shape matrix of category j. It can be shown with relative ease that Eq. 5 in conjunction with Eq. 4 satisfy properties (ii) and (iv), while property (i) is automatically satisfied by the assumed range of T. From these two equations we readily obtain

$$T(j\,|\,\mathbf{x}) = \begin{cases} 1 - \dfrac{q_j(\mathbf{x}) + \sqrt{q_j^2(\mathbf{x}) - (R_j'^2 - d_{mc}^2)d_{cx}^2}}{R_j'^2 - d_{mc}^2} & R'_j > d_{mx} \\ 0 & d_{mx} \ge R'_j \end{cases} \tag{6}$$

where we have defined the following quantities:

$$q_j(\mathbf{x}) \doteq (\mathbf{m}_j - \mathbf{c}_j)^T \mathbf{C}_j (\mathbf{c}_j - \mathbf{x}) \tag{7}$$

$$R'_j \doteq R_j + \delta \tag{8}$$

$$d_{mc} \doteq \|\mathbf{m}_j - \mathbf{c}_j\|_{C_j} \tag{9}$$

$$d_{cx} \doteq \|\mathbf{c}_j - \mathbf{x}\|_{C_j} \tag{10}$$

$$d_{mx} \doteq \|\mathbf{m}_j - \mathbf{x}\|_{C_j} \tag{11}$$

Moreover, it is straightforward to observe that the CCF in Eq. 6 also satisfies condition (iii). The derived CCF for FASEAM depends on the relative distances between the training pattern, the center, and centroid, as measured in the category's metric (expressed by its shape matrix), as well as on the category's Mahalanobis radius. Figure 4 shows a contour plot of the derived CCF for a typical FASEAM category in a 2-

dimensional setting, which verifies that the CCF given in Eq. 6 indeed satisfies conditions (i) through (iv).

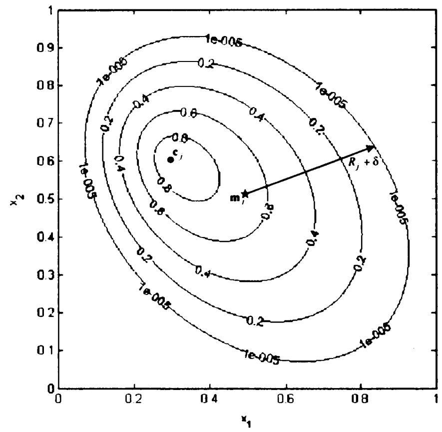

Fig. 4. Contour plot of the FASEAM CCF for an arbitrary category assuming a 2-dimensional input space.

Additionally, we had to replace the match tracking mechanism with an alternative secondary search process, since match tracking in FASART is mostly ineffective: during fast learning, when a winning category learns a pattern after match tracking has been invoked, if we were to immediately present again the same training pattern, the pattern may access a completely different category. In other words, FASART with match tracking does not preserve the principle of instantaneous, incremental learning of FAM, a fact that was first pointed out in [11]. To that effect, when a winning category fails the prediction test for a specific pattern, instead of invoking match tracking, FASEAM was designed to leave the vigilance parameter value unchanged and search for a suitable category that passes the prediction test and, if updated, to preserve the aforementioned principle.

IV. EXPERIMENTAL RESULTS

In order to assess the effectiveness of our design choices regarding FASEAM, we conducted a series of experiments using four artificially-generated, classification data sets. Based on these sets we compared FASART, EAM and FASEAM classifiers in terms of generalization performance and post-training structural complexity. In the next subsection we provide a short description for each data set.

A. Description of data sets

1) 4-Gaussian Datasets: Three datasets were generated by sampling from a bi-variate mixture of isotropic Gaussian distributions with equal priors consisting of 4 components; each component corresponded to a separate class distribution. The means of the class conditional distributions are placed symmetrically with respect to the coordinate axes, so that by changing their relative separation the Bayes error can be analytically calculated. We generated 3 datasets named G4LO, G4ME and G4HI with predetermined Bayes errors 0.05, 0.15 and 0.4 respectively. Each data set consisted of a training set, a cross-validation set and a test set of 500, 5,000 and 5,000 patterns respectively. The training set was kept small to facilitate speedier training phases, while the cross-validation and test set were chosen to be large, so that the statistical tests comparing the models' generalization performance would have good resolution in determining superiority among similarly performing networks.

2) Noisy Circle in the Square: The data set (abbreviated as NCINS) consists of 2-dimensional input data sampled from within the unit square $[0,1]^2$. In the noiseless version of the related classification problem, uniformly sampled data points, that are located within a specified radius r from $[0.5\ 0.5]^T$, are labeled as '1' and the rest as '0'. The radius r is chosen so that both classes have equal prior probabilities. In the noisy version of the problem, the label of the noiseless patterns is flipped with probability $p \leq 0.5$ resulting in a classification problem with Bayes probability of error equal to p. We chose a value of p=0.1. The classifier's task is to learn the decision boundary of the 2-class problem (the circle or radius r centered at $[0.5\ 0.5]^T$) despite the presence of the noisy patterns. Training, cross-validation and test sets were generated with cardinalities 500, 4000 and 4000 respectively.

B. Experimental Setup

For each dataset we trained a large number of FASART, EAM and FASEAM classifiers using several thousands combinations of training parameter values. Let us mention here that the same combinations of parameter values for each classifier type were used for all 4 data sets. Next, for a given data set we identified via cross-validation the 100 best-performing classifiers from each model family, whose generalization performance we subsequently assessed on the test set. With respect to training parameter values, for EAM we used the Weber Law CCF with a constant value for the choice parameter α=0.001 and a CCF value of 0 for uncommitted, F_2 layer nodes. For both EAM and FASEAM the axes ratio μ took values in [0.2:0.1:1.0]. Additionally, for both FASART and FASEAM the δ parameter took values in [0.05:0.05:0.5]. Finally, for FASART the learning rate β_c used for updating the centroid was held at constant value of 0.05.

Settings common to all three architectures were (i) a baseline vigilance $\bar{\rho}$ in [0.0:0.05:0.95] for the training phase, (ii) a baseline vigilance $\bar{\rho}$ of 0 during performance phase to force the classification of all cross-validation and test patterns and (iii) a learning rate β of 1 for min- and max-vector updates (fast learning mode). Also, all three classifier types were trained with 50 different orders of training pattern presentations. The above training parameter values in conjunction with the 50 different presentation orders resulted in 9,000 EAM, 10,000 FASART and 90,000 FASEAM trained networks for each dataset; 436,000 models were trained in total.

C. Observations

In the following presentation and discussion of the results PIC will stand for percent incorrect classification, that is, the error rate of a classifier, while PCC will stand for percent correct classification (equals 100%-PIC). Additionally, we measure the size (structural complexity) of an architecture by the number of categories created during training. Ideally, a classifier should have the lowest possible PIC (equal to the Bayes error) and have the smallest possible size (equal to the number of classes) for a given classification problem. All inter-model comparisons drawn are based on test set performance.

Table I depicts the maximum, median, minimum and standard deviation for the PCC as measured on the test set for each classifier type and each data set considered. Furthermore, in this table, best values for each row are depicted in bold. On the other hand, Figures 5(a), 5(b), 5(c) and 5(d) depict the generalization performance (PIC on test set) versus structural complexity (size) of the 10 best models from each network family; each plot corresponds to one of the four data sets. In the sequel, statements about a difference in PIC (or PCC) being statistically significant are based on a test of hypothesis with significance level (Type I error probability) of 0.01. The test's null hypothesis amounts to the two classifiers compared being equally good (0 difference in PIC/PCC), while the alternative hypothesis is that the classifier featuring the highest PCC (lowest PIC) point estimate indeed performs better than the other one.

TABLE I

	PCC Test for 100 best networks		
	FASART	**EAM**	**FASEAM**
G4LO			
Max	87.4200	88.0000	**88.4600**
Median	85.9000	87.1000	**87.8000**
Min	85.7000	86.8600	**87.6800**
Std.	0.5219	**0.2422**	0.2451
G4ME			
Max	67.1400	69.6200	**73.7400**
Median	65.3300	68.2600	**72.8000**
Min	64.8000	67.8400	**72.6000**
Std.	0.6017	0.3848	**0.3612**
G4HI			
Max	49.2800	52.3800	**56.9800**
Median	48.5400	51.4100	**56.2400**
Min	47.9000	51.0200	**55.7800**
Std.	0.4232	**0.2801**	0.3656
NCINS			
Max	81.2250	**84.4250**	83.4250
Median	80.4875	**83.2375**	82.1625
Min	80.0750	**82.6250**	81.9250
Std.	0.2811	0.4084	**0.2863**

For the G4LO and NCINS datasets (Bayes errors of 0.05 and 0.1 respectively) Table I reflects that the 100 best networks from each family are comparable in classification performance. On the other hand, for harder problems like G4ME the results show that FASEAM performs better by approximately 5% and 7% than EAM and FASART respectively. Similarly, for the G4HI data set, the homologous differences are about 5% and 8% respectively. These last differences in PCC turn out to be statistically significant. In summa, it seems that FASEAM outperforms the standard EAM classifier in hard classification domains by a noticeable difference. This effect may be attributed to the fact that FASEAM uses a CCF that is not constant within category representation regions and allows for better, maybe smoother, approximation of the decision boundaries than in the case of EAM's CCF.

Turning to Figures 5(a) through 5(d) we may have expected an increased size of the FASART models in comparison to FASEAM and EAM due to FASART's match tracking ineffectiveness. However, this doesn't seem to be the case, maybe because FASART uses a different type of category than FASEAM and EAM (hyper-rectangular versus hyper-ellipsoidal representation regions) and, therefore, the model sizes of these families cannot be compared. However, FASART's training took up to three times more list presentations than EAM and FASEAM models (not shown here), which definitely can be attributed to the ineffectiveness of its match tracking process. Yet another observation pertaining to the 4-Gaussians problems is the high variability in network size of the 10 best EAM classifiers, when compared to the 10 best FASEAM and FASART classifiers, a fact that again could be attributed to the special nature of the CCFs the latter ones employ. For the NCINS dataset all three types of models exhibited noticeable variation in network size, which could be due to the problem's uniform class overlap. Nevertheless, FASEAM's performance is statistically indistinguishable from EAM's, while it employs less categories.

V. Summary

In this paper we presented Fuzzy Adaptive System Ellipsoid ARTMAP (FASEAM), a novel neural architecture based on Ellipsoid ARTMAP (EAM) that is designed around the framework utilized in the Fuzzy Adaptive System ART (FASART) architecture. The design was made by augmenting EAM categories with an adjustable centroid vector and utilizing an appropriate category choice function (CCF) that shares the major properties of FASART's CCF. After performing a series of experiments using four artificially-generated data sets, we obtained preliminary indications that FASEAM may perform (in terms of classification) significantly better than the standard EAM architecture, especially when the classification problem features highly overlapping class distributions. This fact could be attributed to FASEAM's CCF that is not constant within category representation regions and allows for better, maybe smoother, approximation of the decision boundaries involved.

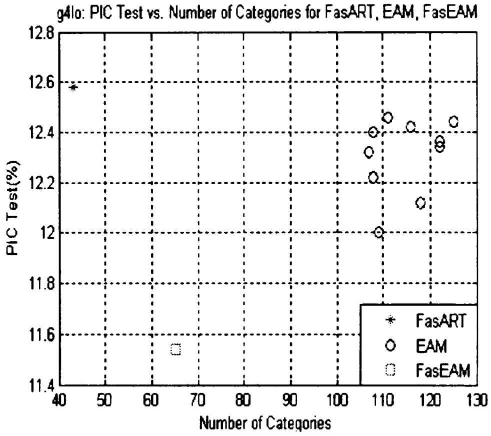

Fig. 5(a). PIC on test set versus network size; 4-G dataset; 5% overlap.

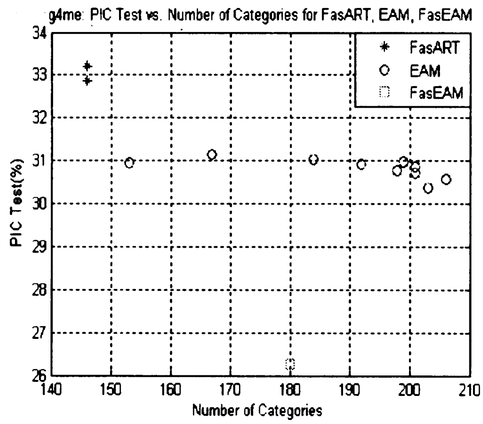

Fig. 5(b). PIC on test set versus network size; 4-G dataset; 15% overlap.

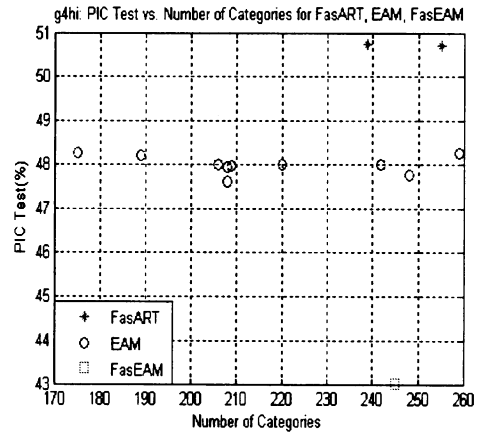

Fig. 5(c). PIC on test set versus network size; 4-G dataset; 40% overlap.

ACKNOLEDGEMENT

The authors would like to thank Manuel Fernández Delgado for helpful comments and suggestions on the manuscript. Also, the authors' research was partially supported by the National Science Foundation under CCLI-EMD grant No. 0341601.

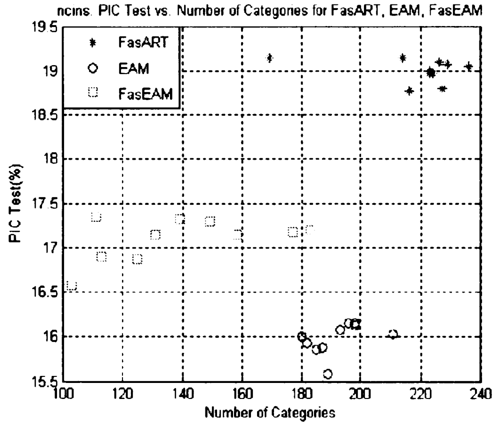

Fig. 5(d). PIC on test set versus network size; NCINS dataset; 10% overlap.

REFERENCES

[1] S. Grossberg, "Adaptive Pattern Recognition and Universal Encoding II: Feedback, Expectation, Olfaction, and Illusions," Biological Cybernetics, 23, pp. 187-202, 1976.

[2] Cano, J. M., Dimitriadis, Y. A., Arau¬zo, M. J., & Coronado, J. "FasArt: A new neuro-fuzzy architecture for incremental learning in systems identification," *Proceedings of 13th World Congress of IFAC*, San Francisco, 1996, vol. F, pp. 133-138.

[3] 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 Transaction on Neural Networks, Vol. 3:5, pp. 698-713, 1992.

[4] Gomez, E., Gago, J. A., Dimitriadis, Y. A., Cano, J. M., & Coronado, J., "Experimental study of a novel neuro-fuzzy system for on-line handwritten UNIPEN digit recognition," Pattern Recognition Letters, 19 (3), pp. 357-364, 1998.

[5] Sainz, G. I., Dimitriadis, Y. A., Cano, J. M., Go¬mez, E., & Parrado, E. *ART based model set for pattern recognition: FasArt family*. In H. Bunke, & A. Kandel (Eds.), Neuro-fuzzy pattern recognition. World Scientific Publ. Co., 2000.

[6] Anagnostopoulos, G.C., & Georgiopoulos, M. "Category Regions as New Geometrical Concepts in Fuzzy ART and Fuzzy ARTMAP," Neural Networks, 15(10), 1205-1221, 2002.

[7] Cano-Izquierdo, J.M., Dimitriadis, Y.A., Gomez-Sanchez, E. And Coronado, J.L., "Learning from noisy information in FasArt and FasBack neuro-fuzzy systems," Neural Networks, 14(4,5), pp. 407-425, 2001.

[8] Anagnostopoulos, G.C., & Georgiopoulos, M. "Ellipsoid ART and ARTMAP for Incremental Unsupervised and Supervised Learning," *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN '01)*, Washington, Washington D.C, 2001, Vol. 2, (pp. 1221-1226).

[9] Gomez-Sanchez, E., "ART neural architectures for pattern recognition and function approximation", Ph.D. Dissertation, University of Valladolid, Spain, 2001.

[10] Williamson, J.R., "Gaussian ARTMAP: A Neural Network for Fast Incremental Learning of Noisy Multidimensional Maps", Neural Networks, 16(1), pp. 881-897, 1996.

[11] Parrado-Hernandez, E., Gomez-Sanchez, E. and Dimitriadis, Y.A., "Study of distributed learning as a solution to category proliferation in Fuzzy ARTMAP based neural systems", Neural Networks, 16(7), pp. 1039-1057, 2003.