COMPARISON OF CLASSIFIERS FOR RADAR EMITTER TYPE IDENTIFICATION

ERIC GRANGER[1,2], STEPHEN GROSSBERG[3], PIERRE LAVOIE[1] and MARK A. RUBIN[3]

[1] Defence Research Establishment Ottawa
Department of National Defence, Ottawa, Canada, K1A 0Z4

[2] Department of Electrical and Computer Engineering
École Polytechnique de Montréal, Montreal, Canada, H3C 3A7

[3] Department of Cognitive and Neural Systems
Boston University, Boston, MA 02215, USA

ABSTRACT

ARTMAP neural network classifiers are considered for the identification of radar emitter types from their waveform parameters. These classifiers can represent radar emitter type classes with one or more prototypes, perform on-line incremental learning to account for novelty encountered in the field, and process radar pulse streams at high speed, making them attractive for real-time applications such as electronic support measures (ESM). The performance of four ARTMAP variants – ART-EMAP (Stage 1), ARTMAP-IC, fuzzy ARTMAP and Gaussian ARTMAP – is assessed with radar data gathered in the field. Simulation results indicate that fuzzy ARTMAP and Gaussian ARTMAP achieve an average classification rate consistently higher than that of the other ARTMAP classifiers, and comparable to that of the reference k nearest neighbor (kNN) and radial basis function (RBF) classifiers. ART-EMAP, ARTMAP-IC and fuzzy ARTMAP require fewer training epochs than Gaussian ARTMAP and RBF, and substantially fewer prototype vectors (thus, smaller physical memory requirements and faster fielded performance) than Gaussian ARTMAP, RBF and kNN. Overall, fuzzy ARTMAP performs at least as well as the other classifiers in both accuracy and computational complexity, and better than each of them in at least one of these aspects of performance. Incorporation into fuzzy ARTMAP of the MT- feature of ARTMAP-IC is found to be essential for convergence during on-line training with this data set.

1. INTRODUCTION

Radar electronic support measures (ESM) involve the search for, interception, location, analysis and identification of radiated electromagnetic energy for military purposes. This provides valuable information for real-time threat detection, threat avoidance, and for timely deployment of counter-measures (Davies and Hollands, 1982) (Wiley, 1993).

A critical function of radar ESM is the real-time identification of the radar type associated with each pulse train, or continuous wave signal, that is intercepted. This task is exceedingly challenging owing to increases in environment density (e.g., pulse Doppler radars can transmit hundreds of thousands of pulses per second); dynamically changing environments; multiplication and dispersion of the modes for military radars; agility in parameters like pulse repetition interval, radio frequency and scan; unknown and reserve modes for which no ESM library entry exists; overlap between the parameters of different radar types in the ESM library; and noise and propagation effects that lead to erroneous or incomplete signal characterization. New classification algorithms to perform radar type identification are thus sought.

In this paper, artificial neural networks are examined using computer simulations. Since artificial neural network classifiers are trained on samples of the data that they are intended to identify, radar data collected in the field were used in the simulations. From an ESM standpoint, training a system directly on radar data is a radical departure from current practice. At present, data are collected, analyzed, and distilled into ESM libraries off-line by skilled analysts. New libraries are disseminated to the field as needed. One inconvenience of the current approach is that it is time-consuming and does not allow rapid learning of new radar modes in the field. The neural network approach may thus constitute an interesting addition to radar ESM.

An attractive feature of ARTMAP (Carpenter et al., 1991a) neural network classifiers is that new classes can be learned incrementally without retraining on the whole data set. ARTMAP networks can also learn one or more prototypes to represent a radar type class, lend themselves well to high speed processing of radar pulses, and can perform familiarity discrimination to avoid meaningless guesses on unfamiliar types (Granger et al., 1999). Four networks based on ARTMAP are compared in this paper. They are ART-EMAP (Stage 1) (Carpenter and Ross, 1995), ARTMAP-IC (Carpenter and Markuzon, 1998), fuzzy ARTMAP (Carpenter et al., 1992) and Gaussian ARTMAP (Williamson, 1996) (Williamson, 1997). The k nearest neighbor (kNN) (Cover and Hart, 1967) and radial basis function (RBF) (Chen et al., 1991) classifiers are included for reference. Parameters used for classification are radio frequency, pulse repetition interval, and pulse width. The performance of the classifiers is assessed in terms of classification rate and computational complexity.

The ARTMAP neural network classifier and the four variants in this study are briefly reviewed in the next section. The radar data set used to train and test the classifiers is presented in Section 3. Simulation results are then presented and analyzed in Section 4.

2. ARTMAP NEURAL NETWORK CLASSIFIERS

ARTMAP refers to a family of neural network architectures for self-organizing recognition and prediction (Carpenter et al., 1991a) (Carpenter et al., 1992). A simplified ARTMAP architecture is built by combining an ART (Carpenter and Grossberg, 1987) neural network with a map field. The ART network consists of two fully connected layers of nodes: an M' node input layer, $F1$, and an N node competitive layer, $F2$. A set of real weights $\mathbf{W} = \{w_{ji} : j = 1,2,...,N; i = 1,2,...,M'\}$ is associated with the $F1 - F2$ layer connections. Each $F2$ node j represents a category in the input space, and stores a prototype vector $\mathbf{w}_j = (w_{j1}, w_{j2}, ..., w_{jM'})$. The $F2$ layer is connected, through associative links, to an L node map field F^{ab}, where L is the number of classes in the output space. A set of binary weights $\mathbf{W}^{ab} = \{w^{ab}_{jk} : j = 1,2,...,N; k = 1,2,...,L\}$ is associated with the $F2 - F^{ab}$ connections. The vector $\mathbf{w}^{ab}_j = (w^{ab}_{j1}, w^{ab}_{j2}, ..., w^{ab}_{jL})$ relates $F2$ node j to one of the L output classes.

During training, ARTMAP classifiers perform incremental supervised learning of the mapping between training set patterns $\mathbf{a} = (a_1, a_2, ..., a_M)$ and respective supervision patterns $\mathbf{b} = (b_1, b_2, ..., b_L)$. The binary supervision patterns are coded to have unit value $b_K = 1$, where K is the target class label for $\mathbf{a}$, and zero elsewhere. The following algorithm describes the operation of an ARTMAP classifier in learning mode:

1. **Initialization:** Initially, all the neurons of $F2$ are uncommitted, all weight values w_{ji} are initialized to 1, and all weight values w^{ab}_{jk} are set to 0. Values of parameters such as the baseline vigilance $\bar{\rho} \in [0,1]$ are also set.
2. **Input pattern coding:** When a training pair $(\mathbf{a}, \mathbf{b})$ is presented to the network, $\mathbf{a}$ undergoes preprocessing, and yields pattern $\mathbf{A} = (A_1, A_2, ..., A_{M'})$. The vigilance parameter ρ is reset to its baseline value $\bar{\rho}$.
3. **Prototype selection:** Pattern $\mathbf{A}$ activates layer $F1$ and is propagated through weighted connections $\mathbf{W}$ to layer $F2$. Activation of each node j in the $F2$ layer is determined by the *choice function*. The $F2$ layer produces a binary, winner-take-all pattern of activity $\mathbf{y} = (y_1, y_2, ..., y_N)$ such that only node $j = J$ with the greatest activation value remains active ($y_J = 1$). Node J propagates its prototype vector $\mathbf{w}_J$ back onto

$F1$ and the *vigilance test* is performed. This test compares the degree of match between $\mathbf{w}_J$ and $\mathbf{A}$ to the vigilance parameter $\rho \in [0,1]$. If this test is satisfied, node J remains active and resonance is said to occur. Otherwise, the network inhibits the active $F2$ node and searches for another node J that passes the vigilance test. If such a node does not exist, an uncommitted $F2$ node becomes active and undergoes learning (Step 5).

4. **Class prediction:** Pattern $\mathbf{b}$ is fed directly to the map field F^{ab}, while the $F2$ activity pattern $\mathbf{y}$ is propagated to the map field via associative connections $\mathbf{W}^{ab}$. The latter input activates F^{ab} nodes according to the *prediction function*, and the most active F^{ab} node K yields the class prediction ($K = k(J)$). If node K constitutes an incorrect class prediction, a *match tracking* signal raises vigilance just enough to induce another search among $F2$ nodes (Step 3). This search continues until either an uncommitted $F2$ node becomes active (learning ensues at Step 5), or a node J that has previously learned the correct class prediction K becomes active.
5. **Learning:** Learning input $\mathbf{a}$ involves updating prototype vector $\mathbf{w}_J$, and, if J corresponds to a newly-committed node, creating a permanent associative link to F^{ab}. A new association between $F2$ node J and F^{ab} node K ($k(J) = K$) is learned by setting $w^{ab}_{Jk} = 1$ for $k = K$, where K is the target class label for $\mathbf{a}$.

Once the weights ($\mathbf{W}$ and $\mathbf{W}^{ab}$) have converged for the training set patterns, ARTMAP can predict a class label for an input pattern by performing Steps 2, 3 and 4 without any testing. A pattern $\mathbf{a}$ that activates node J is predicted to belong to the class $K = k(J)$.

Although the first ARTMAP (Carpenter et al., 1991a) classifier is limited to processing binary-valued input patters, the ART-EMAP, ARTMAP-IC, fuzzy ARTMAP (FAM) and Gaussian ARTMAP (GAM) classifiers can process both analog and binary-valued input patterns. The rest of this section highlights the main differences between these four ARTMAP variants. (Refer to Table 1 for algorithmic differences.)

2.1 ART-EMAP, ARTMAP-IC and fuzzy ARTMAP

ART-EMAP, ARTMAP-IC and FAM variants employ fuzzy ART (Carpenter et al., 1991b) as the ART network. With fuzzy ART, a transformation called complement coding doubles the number of components in the input pattern ($M' = 2M$), which becomes $\mathbf{A}$ = $(\mathbf{a}, \mathbf{a}^c)$ = $(a_1, a_2, ..., a_M; a^c_1, a^c_2, ..., a^c_M)$, where $a^c_i = (1 - a_i)$, and $a_i \in [0,1]$. With complement coding and fast learning ($\beta = 1$), fuzzy ART represents category j as hyperrectangle R_j that just encloses all the training set patterns $\mathbf{a}$ to which it has been assigned. That is, a $2M$-dimensional prototype vector $\mathbf{w}_j = (w_{j1}, w_{j2}, ..., w_{j2M})$ records the largest and smallest component values of training set patterns $\mathbf{a}$ placed in the j^{th} category.

ART-EMAP (Stage 1) and ARTMAP-IC are extensions of FAM that produce a binary winner-take-all pattern $\mathbf{y}$ when training, but use distributed activation of coded $F2$ nodes when testing. ARTMAP-IC is further extended in two ways. First, it biases distributed test set predictions according to the number of times $F2$ nodes are assigned to training set patterns. Second, it uses negative match tracking (i.e., negative ε values), denoted MT-, to address the problem of inconsistent cases, whereby identical training set patterns correspond to different classes labels.

2.2 Gaussian ARTMAP

GAM differs significantly from the other three. It represents category j as a Gaussian density function, defined by two vectors: its mean $\mu_j = (\mu_{j1}, \mu_{j2}, ..., \mu_{jM})$ and its standard deviation $\sigma_j = (\sigma_{j1}, \sigma_{j2}, ..., \sigma_{jM})$ (together they replace prototype vector $\mathbf{w}_j$). A scalar, n_j, accumulates the amount of relative activation obtained by $F2$ node j on training set patterns.

During training, the number of committed $F2$ nodes, N_c, is initially set to 0. Newly-committed $F2$ nodes increment N_c, and undergo the initialization step (instead of the learning step). The committed $F2$ nodes that pass the vigilance test for pattern $\mathbf{a}$ are allowed to

Table 1: Distinctive equations used by the ARTMAP networks. With ART-EMAP, ARTMAP-IC and FAM, $\lvert\cdot\rvert$ is the norm operator ($\lvert\mathbf{w}_j\rvert \equiv \sum_{i=1}^{2M} \lvert w_{ji}\rvert$), $\wedge$ is the fuzzy AND operator ($(\mathbf{A}\wedge\mathbf{w}_j)_i \equiv \min(A_i, w_{ji})$), α is the *choice parameter*, and β is the *learning rate parameter*. With GAM, γ is the initial standard deviation assigned to newly-committed $F2$ nodes.

Algorithmic step	**ARTMAP classifier**	
	ART-EMAP, ARTMAP-IC and FAM	GAM
1. Initialization:	$\alpha > 0$, $\beta \in [0,1]$	$\mu_J = \mathbf{A}$, $\sigma_{Ji} = \gamma$ $(\gamma > 0)$, $w^{ab}_{JK} = 1$, $n_J = 1$
2. Input pattern coding:	$\mathbf{A} = (\mathbf{a}, \mathbf{a}^c)$ $(M' = 2M)$	$\mathbf{A} = \mathbf{a}$ $(M' = M)$
3. Prototype selection:		
– choice function	$T_j(\mathbf{A}) = \lvert\mathbf{A}\wedge\mathbf{w}_j\rvert/(\alpha + \lvert\mathbf{w}_j\rvert)$	$g_j(\mathbf{A}) = \begin{cases} \frac{n_j}{\prod_{i=1}^{M}\sigma_{ji}} G_j(\mathbf{A}) & \text{if } G_j(\mathbf{A}) > \rho \\ 0 & \text{otherwise} \end{cases}$
– vigilance test	$\lvert\mathbf{A}\wedge\mathbf{w}_J\rvert \geq \rho M$	$G_j(\mathbf{A}) = \exp\left\{-\frac{1}{2}\sum_{i=1}^{M}\frac{(A_i-\mu_{ji})^2}{\sigma_{ji}^2}\right\} > \rho$
– $F2$ activation	$y_j = 1$ only if $j = J$ (winning node)	$y_j = g_j/(0.01 + \sum_{l=1}^{N_c} g_l)$
4. Class prediction:		
– prediction function	$S^{ab}_k(\mathbf{y}) = \sum_{j=1}^{N} y_j w^{ab}_{jk}$	$S^{ab}_k(\mathbf{y}) = \sum_{j=1}^{N_c} y_j w^{ab}_{jk}$
– match tracking	$\rho' = (\lvert\mathbf{A}\wedge\mathbf{w}_J\rvert/M) + \varepsilon$	$\rho' = \exp\left\{-\frac{1}{2}\sum_{j\in E_K} y_j^* \sum_{i=1}^{M}\frac{(A_i-\mu_{ji})^2}{\sigma_{ji}^2}\right\} + \varepsilon$
5. Learning:		
– prototype update	$\mathbf{w}'_J = \beta(\mathbf{A}\wedge\mathbf{w}_J) + (1-\beta)\mathbf{w}_J$	$n'_j = n_j + y^*_j$ $\mu'_{ji} = (1 - \frac{y^*_j}{n_j})\mu_{ji} + \frac{y^*_j}{n_j} A_i$ $\sigma'_{ji} = \sqrt{(1 - \frac{y^*_j}{n_j})\sigma^2_{ji} + \frac{y^*_j}{n_j}(A_i - \mu_{ji})^2}$

activate, and distribute a pattern of activity $\mathbf{y} = (y_1, y_2, ..., y_{N_c})$. Match tracking and learning are performed according to the relative activation over the "ensemble" E_K of $F2$ nodes linked to the predicted F^{ab} node K. The relative activation over E_K is defined by the distributed pattern $\mathbf{y}^* = (y_1^*, y_2^*, ..., y_{N_c}^*)$, where $y_j^* = y_j / \sum_{l \in E_K} y_l$ only if $j \in E_K$, and $y_j^* = 0$ otherwise.

3. RADAR PULSE DATA

The data set used to evaluate the classifiers contains approximately 100,000 consecutive radar pulses gathered over 16 seconds during a field trial by the Defense Research Establishment Ottawa. After the trial, an ESM analyst manually separated trains of pulses coming from different emitters. Each pulse was then tagged with two numbers: a radar type number and a mode number. A single type of radar can use several modes to perform various functions. Since ESM trials are complex and never totally controlled, not all pulses can be tagged and a residue is obtained. Residue pulses were discarded for this study.

The patterns used for this study contain 3 parameters: radio frequency (RF), pulse repetition interval (PRI) and pulse width (PW). The RF and PW parameters are measured on each individual pulse, whereas PRI is derived from the time-of-arrival (TOA) of pulses from a single emitter. For simplicity, a TOA deinterleaver was assumed to be capable of correctly sorting the n_k pulses belonging to each active emitter mode k, and then computing: $\mathrm{PRI}_k(i) = \mathrm{TOA}_k(i) - \mathrm{TOA}_k(i-1)$ for $i = 2, 3, ..., n_k$. Note that the first pattern from each active emitter mode is omitted from the comparison. Also, due to the circular scanning action of some radar emitters, pulses are recorded in bursts. The first pulse of each scan (or burst) is also omitted. Finally, the parameters were linearly normalized so that $a_i \in [0, 1]$, $i = 1, 2, 3$.

Once deinterleaved and tagged, the data used to train and test the classifiers contain 52,192 radar pulses from 34 modes, each one belonging to one of 15 different radar types. The data feature bursts of high pulse densities, multiple emitters of a same type, modes with overlapping parametric ranges, radars transmitting at different pulse rates, and emitters switching modes. The sophistication of the radar types range from simple (constant RF and PRI) to fairly complex (pulse-to-pulse RF and PRI agility). The data also contain direction of arrival (DOA), but this parameter is not used here.

4. SIMULATION RESULTS AND DISCUSSION

Matlab code was written for all the classifiers. Prior to each simulation trial, the data set described in Section 3 was partitioned into training and test subsets. 50% of the data from each radar type was selected at random to form the training subset. Then, training subset patterns **a** were repeatedly presented to each classifier along with their supervision patterns **b** (coding radar types) until convergence was reached – when the sum-squared-fractional-change (SSFC) of prototype weights was less that 0.001 for two successive epochs[1]. An epoch is defined as a presentation of the training subset to a classifier in a TOA sequence. Finally, the test subset (the complete set less the training data) was presented to the trained classifiers for prediction.

The performance of the classifiers was compared in terms of the amount of resources they require during training, and their predictive accuracy on the test subset. The amount of resources required while training is measured in the 3 following ways. *Compression* refers to the average ratio of training patterns to committed F2 node. *Memory* is the number of normalized registers[2] needed to store the set of learned prototype vectors. The *convergence*

[1] The RBF classifier used (Chen et al., 1991) selects training subset patterns one by one to create hidden (RB) layer nodes. Convergence is said to have been reached when the sum-squared-error between actual outputs (resulting from training patterns) and target outputs falls below 0.01.

[2] A normalized register is a fixed-size register, with the number of bits equal to that needed to store the classifier's real values such as a_i, w_{ji}, ρ, etc.

time is the number of epochs required for the classifier to converge (SSFC < 0.001). Predictive accuracy is measured with the *classification rate* – the ratio of correctly classified patterns over all test patterns.

Table 2 presents the average results (with corresponding standard deviations) of each classifier, from 20 independent simulation trials. The parameter settings were selected to achieve the best classification rate for the least amount of memory and convergence time during training.

Results in Table 2 indicate that FAM and GAM yield the highest average classification rates for the data set, followed by ARTMAP-IC and ART-EMAP. These classification rates are comparable to those obtained with the reference kNN and RBF classifiers. ART-EMAP, ARTMAP-IC and FAM attain their classification rates with greater compression (thus less memory requirement to store prototype vectors) than the other classifiers, and significantly fewer training epochs to converge than GAM and RBF.

4.1 Training set convergence

A convergence problem occurs with ART-EMAP, FAM and GAM whenever the training subset contains identical patterns (i.e., pulses in the same resolution cell) that belong to different radar types. The problem is aggravated since ARTMAP tends to segment the overlapping/scattered parts of classes into several very tiny, often minimum-sized prototypes. The effect is a proliferation of identical prototypes for certain training set patterns.

Consider the following example with FAM. Assume that on the first training epoch, FAM learns two completely overlapping, minimum-sized prototypes, $\mathbf{w}_{A.1}$ (linked to class A) and $\mathbf{w}_{B.1}$ (linked to class B), for two identical pulse patterns, $\mathbf{a}_1$ and $\mathbf{a}_2$. In a subsequent epoch, $\mathbf{w}_{A.1}$ is initially selected to learn $\mathbf{a}_2$, since $T_{A.1} = T_{B.1}$, and $\mathbf{w}_{A.1}$ was created prior to $\mathbf{w}_{B.1}$ (index $A.1$ is smaller than $B.1$). Since $\mathbf{w}_{A.1}$ is not linked to class B, mismatch reset raises the vigilance parameter ρ to $(|\mathbf{A}_2 \wedge \mathbf{w}_{A.1}|/M) + \varepsilon$, where $|\mathbf{A}_2 \wedge \mathbf{w}_{A.1}| = |\mathbf{A}_2 \wedge \mathbf{w}_{B.1}|$. As a result, $\mathbf{w}_{B.1}$ can no longer pass the vigilance test required to become selected for $\mathbf{a}_2$, and FAM must create another minimum-sized prototype $\mathbf{w}_{B.2} = \mathbf{w}_{B.1}$. From epoch to epoch, the same phenomenon repeats itself, yielding ever more prototypes $\mathbf{w}_{B.\mathrm{n}} = \mathbf{w}_{B.1}$ for $\mathrm{n} = 3, 4, \ldots, \infty$.

ART-EMAP, FAM and GAM failed to converge while training on the radar data set. Results in Table 2 were obtained through manual termination by: (1) detecting, from epoch to epoch, the repeated creation of identical prototypes for the same training pulses, (2) pruning non-unique prototypes from memory, and (3) defining the convergence time as the number of epochs leading to the creation of non-duplicate prototypes only. Clearly this approach is cumbersome, which is why a better way of handling the problem was examined.

ARTMAP-IC was capable of converging incrementally on-line for the radar data. The feature of ARTMAP-IC that allows it to avoid the convergence problem is MT-; that is, the use of a negative ε value (Carpenter and Markuzon, 1998). In the example above, mismatch reset would raise ρ but $\mathbf{w}_{B.1}$ would still pass the vigilance test. This allows ARTMAP-IC to learn fully overlapping prototypes for training set patterns that belong to different classes. As is seen from Table 2, and as is discussed further in Subsection 4.4, the complete ARTMAP-IC algorithm – MT-, distributed activation and instance counting – performs comparatively poorly on this data. This suggests that the best approach is ARTMAP-IC with MT- only, or, equivalently, FAM plus MT-.

4.2 Distributing and biasing test set activation

Simulation trials showed that the Q-max rule (Carpenter and Markuzon, 1998) for distributing $F2$ layer activation in ART-EMAP and ARTMAP-IC classifiers gives better results than either power or threshold rules (Carpenter and Ross, 1995) with the radar data. However, a priori choice of the Q parameter for this rule is critical for performance. The following choice of Q was found to give good results for the radar data: $Q = \min\{\lceil N_c/2L \rceil, 2L\}$, where L is the number of classes (15 in this study) and N_c is the number of committed $F2$

Table 2: Average classification results for the radar data. The * indicates that the classifier was unable to converge for the training set on each trial. (Refer to Subsection 4.1 for further details.)

Classifier	**Evaluation criteria (std. deviation)**			
	Accuracy	**Resources**		
	Classification rate	**Compression**	**Memory**	**Convergence time**
kNN ($k = 1$, $d_{\text{cityblock}}$)	0.996 (0.001)	1 (0)	80311 (0)	N/A
kNN ($k = 1$, $d_{\text{Euclidean}}$)	0.996 (0.001)	1 (0)	80311 (0)	N/A
RBF* (spread of RB node = 0.05)	0.997 (0.001)	4.1 (0.1)	6367.9 (65.1)	2123.6 (21.7)
ART-EMAP* (Stage 1) ($\varepsilon = 10^{-4}$, $\alpha = .001$, $\beta = 1$, $\bar{\rho} = 0$)	0.784 (0.074)	216.0 (27.1)	735.3 (89.4)	stopped at 3.9 (0.4)
ARTMAP-IC (complete) ($\varepsilon = 10^{-4}$, $\alpha = .001$, $\beta = 1$, $\bar{\rho} = 0$)	0.840 (0.039)	217.2 (20.5)	727.5 (66.4)	3.8 (0.4)
FAM* ($\varepsilon = 10^{-4}$, $\alpha = .001$, $\beta = 1$, $\bar{\rho} = 0$)	0.996 (0.001)	216.0 (27.1)	735.3 (89.4)	stopped at 3.9 (0.4)
GAM* ($\varepsilon = 10^{-3}$, $\gamma = .0025$, $\bar{\rho} = 0$)	0.997 (0.001)	102.1 (4.3)	1536.0 (65.8)	stopped at 6.0 (1.1)
FAM with MT- ($\varepsilon = 10^{-6}$, $\alpha = .001$, $\beta = 1$, $\bar{\rho} = 0$)	0.996 (0.001)	211.8 (32.1)	751.8 (90.6)	3.7 (0.5)
FAM with MT- and limited IC ($\varepsilon = 10^{-4}$, $\alpha = .001$, $\beta = 1$, $\bar{\rho} = 0$)	0.996 (0.001)	217.0 (22.5)	727.5 (66.4)	3.8 (0.4)

nodes. In particular, bounding Q to the low maximal value of $2L$ appears to reduce wild performance fluctuations.

Regardless of distributed activation, FAM performs better than its two extensions, ART-EMAP and ARTMAP-IC, on the radar data. Careful observation revealed that distributed activation for the test patterns makes more prediction errors because some radar type classes in the data set are very dispersed, are fragmented, and overlap one another. This is incompatible with class predictions that are based on the distribution of strongly activated $F2$ nodes among radar type classes (rather than on the most active $F2$ node). Indeed, an F^{ab} class node k that receives a very strong activation from one $F2$ node may have weaker overall activation than an F^{ab} class node h that receives a moderately strong activation from several $F2$ nodes. Perhaps this explains why kNN gives its best performance for $k = 1$, and degrades slowly as k grows. (For instance, its classification rate is 0.992 for $k = 9$ and $d_{\text{cityblock}}$.)

GAM involves training and testing with a distributed pattern of activity. $F2$ layer activation is distributed among nodes that pass the vigilance test. When training, each category $j \in E_K$ learns according to its relative activation for $\mathbf{a}$. $F2$ nodes learn a Gaussian mixture model of the input space. Although computationally intensive, the m-learning strategy allows GAM to achieve very high classification rates.

ARTMAP-IC and GAM accumulate weighting factors that depend on the quantity of training subset patterns assigned to each $F2$ node. This frequency information is used to bias predictions towards classes assigned the most training patterns. This constitutes an issue in radar ESM since some critical radars transmit very few pulses, while others transmit hundreds of thousands of pulses per second. Biasing prototype choices according to patterns in the time of arrival of pulses would be more appropriate.

4.3 Prototype representations

Given the quantization of parameter measurements, intercepted radar pulses fall into resolution cells. Since the measurement uncertainty of the three parameters used (RF, PRI and PW) are independent of each other, the radar type definitions are essentially rectangular. ART-EMAP, ARTMAP-IC and FAM use hyperrectangles to represent prototypes in the input space, and appear to be a better match for this type of data. GAM and RBF, on the other hand, represent prototypes with Gaussian density functions. This results in substantial fragmentation of radar type classes, and in low compression for these two classifiers.

4.4 Enhancements to fuzzy ARTMAP

Results obtained with two modified versions of FAM are shown at the bottom of Table 2. The first version augments FAM with the MT- component of ARTMAP-IC. This results in a version of FAM that learns non-unique prototypes linked to different classes (thus converges for the radar data). As shown in Table 2, FAM with MT- performs as well, to within standard error, as kNN, with much less use of memory and without the convergence problem. As pointed out in (Carpenter and Markuzon, 1998) MT- is a better algorithmic approximation to the continuous-time version of the FAM neural network.

Although weighting predictions according to the number of training set patterns is not appropriate in our context, it may be useful on other data sets for FAM with MT- to produce better than a random guess for test patterns that activate the same $F2$ node[3]. This could result from using instance-weighted outputs only in the case where winning nodes are "inconsistent-case siblings," since in this case the only basis on which to choose one of the winning nodes over another may be the frequency with which they were winning nodes during training. Table 2 presents the results for FAM with MT- and "limited instance counting." With this variant, instances are still counted for all $F2$ nodes. However, it distributes activation weighted by the instance counts if and only if nodes J are a set that code

[3]FAM with MT- breaks ties by choosing the $F2$ node with the smallest index.

for the same test pattern but map to different classes. Limited IC does not harm accuracy on the radar data set.

5. CONCLUSIONS

Four neural network classifiers – ART-EMAP (Stage 1), ARTMAP-IC, fuzzy ARTMAP and Gaussian ARTMAP – have been compared for the identification of radar emitter types associated with intercepted radar pulse trains. Their performance has been measured in terms of resource allocation and accuracy.

In computer simulations using a radar data set collected in the field, FAM with negative match tracking (MT-) has performed consistently well. It gives one of the best classification rates, yet requires among the least amount of resources (shortest convergence time and least storage for prototypes) among ARTMAP classifiers. The MT- feature allows it to converge on-line despite the presence of training set categories that overlap completely yet belong to different classes.

ACKNOWLEDGMENTS

This research was supported in part by the Defense Advanced Research Projects Agency and the Office of Naval Research ONR N00014-95-1-0409 (S. G. and M. A. R.), the National Science Foundation NSF IRI-97-20333 (S. G.), the Natural Sciences and Engineering Research Council of Canada (E. G.), and the Office of Naval Research ONR N00014-95-1-0657 (S. G.).

REFERENCES

Carpenter, G. A. and Grossberg, S., 1987, "A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine," *Computer, Vision, Graphics and Image Processing*, **37**, 54-115.

Carpenter, G. A., Grossberg, S. and Reynolds, J. H., 1991a, "ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network," *Neural Networks*, **4**, 565-588.

Carpenter, G. A., Grossberg, S. and Rosen, D. B., 1991b, "Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System," *Neural Networks*, **4:6**, 759-771.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H. and Rosen, D. B., 1992, "Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps," *IEEE Trans. on Neural Networks*, **3:5**, 698-713.

Carpenter, G. A. and Ross, W. D., 1995, "ART-EMAP: A Neural Network Architecture for Object Recognition by Evidence Accumulation," *IEEE Trans. on Neural Networks*, **6:4**, 805-818.

Carpenter, G. A. and Markuzon, N., 1998, "ARTMAP-IC and Medical Diagnosis: Instance Counting and Inconsistent Cases," *Neural Networks*, **11:2**, 323-336.

Chen, S., Cowan, C. F. N. and Grant, P. M., 1991, "Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks," *IEEE Trans. on Neural Networks*, **2:2**, 302-309.

Cover, T. M. and Hart, P. E., 1967, "Nearest Neighbor Patterns Classification," *IEEE Trans. on Information Theory*, **13:1**, 21-27.

Davies, C. L., and Hollands, P., June 1982, "Automatic Processing for ESM," *Proc. IEE*, **129:3 (F)**, 164-171.

Granger, E., Grossberg, S., Rubin, M. A., and Streilein, W. W., 1999, "Familiarity discrimination of radar pulses," In M. S. Kearns et al., eds., *Advances in Neural Information Processing Systems 11*, Cambridge, MA: MIT Press, 875-881.

Wiley, R. G., 1993, *Electronic Intelligence: The Analysis of Radar Signals*, 2nd ed., Artech House.

Williamson, J. R., 1996, "Gaussian ARTMAP: A Neural Network for Fast Incremental Learning of Noisy Multidimensional Maps," *Neural Networks*, **9:5**, 881-897.

Williamson, J. R., 1997 "A Constructive, Incremental-Learning Neural Network for Mixture Modeling and Classification," *Neural Computation*, **9:7**, 1517-1543.