Threshold Determination for ARTMAP-FD Familiarity Discrimination

Gail A. Carpenter, Mark A. Rubin, William W. Streilein
Department of Cognitive and Neural Systems, Boston University
Boston, MA 02215 USA

Abstract *The ARTMAP-FD neural network performs both identification (placing test patterns in classes encountered during training) and familiarity discrimination (judging whether a test pattern belongs to any of the classes encountered during training). ARTMAP-FD quantifies the familiarity of a test pattern by computing a measure of the degree to which the pattern's components lie within the ranges of values of training patterns grouped in the same cluster. This familiarity measure is compared to a threshold which can be varied to generate a receiver operating characteristic (ROC) curve. Methods for selecting optimal values for the threshold are evaluated. The performance of validation-set methods is compared with that of methods which track the development of the network's discrimination capability during training. The techniques are applied to databases of simulated radar range profiles.*

1 Introduction

The recognition process involves both identification and familiarity discrimination. Consider, for example, a neural network designed to identify aircraft based on their radar reflections and trained on sample reflections from ten types of aircraft $A \ldots J$. After training, the network should correctly classify radar reflections belonging to the familiar classes $A \ldots J$, but it should also abstain from making a meaningless guess when presented with a radar reflection from an object belonging to a different, unfamiliar class. Many neural networks carry out pattern recognition, but most perform identification without estimating whether a test set input belongs to a class that became familiar during training [1].

This paper describes ARTMAP-FD, an extension of fuzzy ARTMAP that performs familiarity discrimination. ARTMAP-FD capabilities are demonstrated on data sets of simulated radar range profiles from aircraft targets, with performance evaluated using receiver operating characteristic (ROC) curves. In these simulations, multiwavelength input vectors can have as many as 2400 components, so the application uses the ARTMAP properties of scalability and fast learning in an essential way.

2 Fuzzy ARTMAP

Fuzzy ARTMAP [2] is a self-organizing neural network for learning, recognition, and prediction (Figure 1). Each input $\mathbf{a}$ learns to predict an output class K. During training, the network creates internal recognition categories, with the number of categories determined on-line by predictive success. Components of the vector $\mathbf{a}$ are scaled so that each $a_i \in [0, 1]$ $(i = 1 \ldots M)$. Complement coding [3] doubles the number of components in the input vector, which becomes $\mathbf{A} \equiv (\mathbf{a}, \mathbf{a}^c)$, where the i^{th} component of $\mathbf{a}^c$ is $a_i^c \equiv (1 - a_i)$. With fast learning, the weight vector $\mathbf{w}_j$ records the largest and smallest component values of input vectors placed in the j^{th} category. The $2M$-dimensional vector $\mathbf{w}_j$ may be visualized as the hyperbox R_j that just encloses all the vectors $\mathbf{a}$ that selected category j during training.

Activation of the coding field F_2 is determined by the Weber law choice function $T_j(\mathbf{A}) = | \mathbf{A} \wedge \mathbf{w}_j | / (\alpha + | \mathbf{w}_j |)$, where $(\mathbf{P} \wedge \mathbf{Q})_i \equiv \min(P_i, Q_i)$ and $| \mathbf{P} | \equiv \sum_{i=1}^{2M} | P_i |$. With winner-take-all coding, the F_2 node J that receives the largest $F_1 \rightarrow F_2$ input T_j becomes active. Node J remains active if it satisfies the matching criterion: $| \mathbf{A} \wedge \mathbf{w}_j | / | \mathbf{A} | = | \mathbf{A} \wedge \mathbf{w}_j | / M > \rho$, where $\rho \in [0, 1]$ is the dimensionless *vigilance parameter*. Otherwise, the network resets the

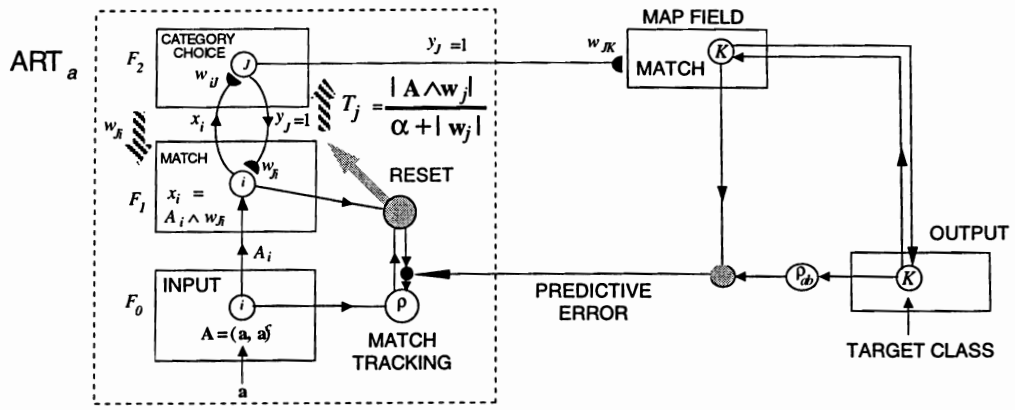

Figure 1: A fuzzy ARTMAP network for classification.

active F_2 node and searches until J satisfies the matching criterion. If node J then makes an incorrect class prediction, a *match tracking* signal raises vigilance just enough to induce a search, which continues until either some F_2 node becomes active for the first time, in which case J learns the correct output class label $k(J) = K$; or a node J that has previously learned to predict K becomes active. During testing, a pattern $\mathbf{a}$ that activates node J is predicted to belong to the class $K = k(J)$.

3 Familiarity discrimination with ARTMAP-FD

3.1 Familiarity measure

During testing, an input pattern $\mathbf{a}$ is defined as *familiar* when a familiarity function $\phi(\mathbf{A})$ is greater than a decision threshold γ. Once a category choice has been made by the winner-take-all rule, fuzzy ARTMAP ignores the size of the input T_J. In contrast, ARTMAP-FD uses T_J to define familiarity, taking

$$\phi(\mathbf{A}) = \frac{T_J(\mathbf{A})}{T_J^{MAX}} = \frac{\mid \mathbf{A} \wedge \mathbf{w}_J \mid}{\mid \mathbf{w}_J \mid}, \tag{1}$$

where $T_J^{MAX} = \mid \mathbf{w}_J \mid /(\alpha + \mid \mathbf{w}_J \mid)$. This maximal value of T_J is attained by each input $\mathbf{a}$ that lies in the hyperbox R_J, since $\mid \mathbf{A} \wedge \mathbf{w}_J \mid = \mid \mathbf{w}_J \mid$ for these points. An input that chooses category J during testing is then assigned the maximum familiarity value 1 if and only if $\mathbf{a}$ lies within R_J.

Note that the choice parameter α in equation (1) is usually taken to be small since the *conservative limit*, where $\alpha = 0^+$, minimizes the number of category nodes formed during training. When $\alpha \approx 0$, $T_J^{MAX} \approx 1$, so $\phi(\mathbf{A}) \approx T_J(\mathbf{A})$. Simulations below set $\alpha = 0.0001$. Then, setting $\phi(\mathbf{A}) = T_J(\mathbf{A})$ produces essentially the same results as setting $\phi(\mathbf{A}) = T_J(\mathbf{A})/T_J^{MAX}$. The former choice of ϕ is more readily computable in a neural network but the latter has a simpler geometric interpretation.

3.2 Familiarity discrimination algorithm

ARTMAP-FD is identical to fuzzy ARTMAP during training. During testing, $\phi(\mathbf{A})$ is computed after fuzzy ARTMAP has yielded a winning node J and a predicted class $K = k(J)$. If $\phi(\mathbf{A}) > \gamma$, ARTMAP-FD predicts class K for the input $\mathbf{a}$. If $\phi(\mathbf{A}) \leq \gamma$, $\mathbf{a}$ is regarded as belonging to an unfamiliar class and the network makes no prediction.

Note that fuzzy ARTMAP can also abstain from classification, when the baseline vigilance parameter $\bar{\rho}$ is greater than zero during testing. Typically $\bar{\rho} = 0$ during training, to maximize code compression. In radar range profile simulations such as those described below, fuzzy ARTMAP can perform familiarity discrimination when $\bar{\rho} > 0$ during both training and testing. However, accurate discrimination requires that $\bar{\rho}$ be close to 1, which causes category proliferation during training.

Range profile simulations have also set $\bar{\rho} = 0$ during both training and testing, but with the familiarity measure set equal to the fuzzy ARTMAP match function:

$$\phi(\mathbf{A}) = \frac{|\mathbf{A} \wedge \mathbf{w}_J|}{M}. \tag{2}$$

This approach is essentially equivalent to taking $\bar{\rho} = 0$ during training and $\bar{\rho} > 0$ during testing, with $\bar{\rho} = \gamma$. However, for a test set input $\mathbf{a} \in R_J$, the function defined by (2) sets $\phi(\mathbf{A}) = |\mathbf{w}_J| / M$, which may be large or small although $\mathbf{a}$ is familiar. Thus this function does not provide as good familiarity discrimination as the one defined by (1), which always sets $\phi(\mathbf{A}) = 1$ when $\mathbf{a} \in R_J$. All the simulations below employ the function (1), with $\bar{\rho} = 0$.

3.3 Familiarity discrimination with sequential evidence accumulation

ART-EMAP (Stage 3) [4] identifies a test set object's class after exposure to a sequence of input patterns, such as differing views, all identified with that one object. Training is identical to that of fuzzy ARTMAP, with winner-take-all coding at F_2. ART-EMAP generally employs distributed F_2 coding during testing. With winner-take-all coding during testing as well as training, ART-EMAP predicts the object's class to be the one selected by the largest number of inputs in the sequence. Extending this approach, ARTMAP-FD accumulates familiarity measures for each predicted class K as the test set sequence is presented. Once the winning class is determined, the object's familiarity is defined as the average accumulated familiarity measure of the predicted class during the test sequence.

4 Familiarity discrimination simulations

Since familiarity discrimination involves placing an input into one of two sets, familiar and unfamiliar, the receiver operating characteristic (ROC) formalism [10,11] can be used to evaluate the effectiveness of ARTMAP-FD on this task. The *hit rate* H is the fraction of familiar targets the network correctly identifies as familiar and the *false alarm rate* F is the fraction of unfamiliar targets the network incorrectly identifies as familiar. Each of these quantities depends upon the decision threshold γ (Section 3.1). An ROC curve is a plot of H *vs.* F, parameterized by γ. The area under the ROC curve is the *c-index*, a measure of predictive accuracy that is independent of both the fraction of positive (familiar) cases in the test set and the positive-case decision threshold γ.

An ARTMAP-FD network was trained on 18 targets from a 36-target set (Figure 2a). Simulations tested sequential evidence accumulation performance for 1, 3, and 100 observations, corresponding to 0.05, 0.15, and 5.0 seconds of observation time (smooth curves, Figure 2b). As in the case of identification [9], a combination of multiwavelength range profiles and sequential evidence accumulation produces good familiarity discrimination, with the c-index approaching 1 as the number of sequential observations grows.

Figure 2b also demonstrates the importance of the proper choice of familiarity measure. The jagged ROC curve was produced by a familiarity discrimination simulation identical to that which resulted in the 100-sequential-view smooth curve, but using the match function (2) instead of ϕ as given by (1).

5 Familiarity threshold selection

The c-index and the shape of the ROC curve indicate a network's potential ability to discriminate between familiar and unfamiliar targets. However, when a system is placed in operation, one particular decision threshold $\gamma = \Gamma$ must be chosen. In a given application, selection of Γ depends upon the relative cost of errors due to missed targets and false alarms. The optimal Γ corresponds to a point on the parameterized ROC curve that is typically close to the upper left-hand corner of the unit square [10,11], to maximize correct selection of familiar targets (H) while minimizing incorrect selection of unfamiliar targets (F).

In the simulations below, three types of hit and false alarm rates are calculated. First, an ROC curve $(F(\gamma), H(\gamma))$ is calculated from the training data. The predicted threshold $\gamma = \Gamma_P$ is then chosen, either by a validation set procedure that seeks to minimize $[F(\gamma) - H(\gamma)]$

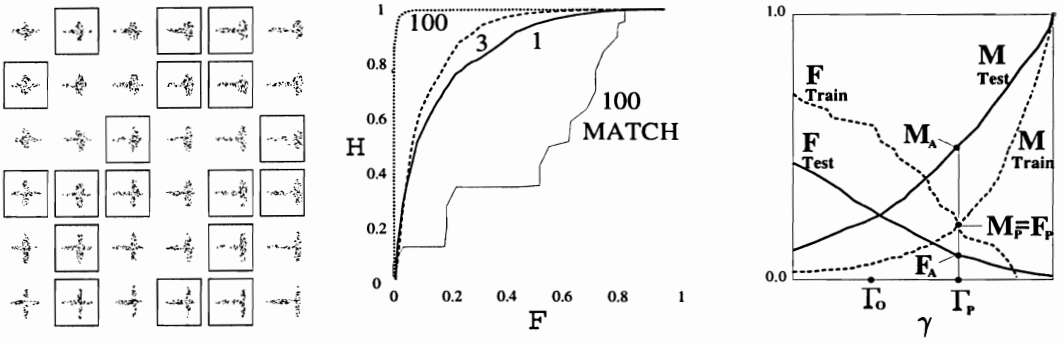

Figure 2: (a) 36 simulation targets with 6 wing positions and 6 wing lengths, and 100 scattering centers per target. Boxes indicate randomly selected familiar targets. (b) ROC curves from ARTMAP-FD simulations, with multiwavelength range profiles having 40 center frequencies. Sequential evidence accumulation for 1, 3 and 100 views uses familiarity measure (1) (smooth curves); and for 100 views uses the match function (2) (jagged curve). (c) Training and test curves of miss rate $M = (1 - H)$ and false alarm rate F *vs* threshold γ, for 36 targets and one view. Training curves intersect at the point where $\gamma = \Gamma_P$ (predicted); and test curves intersect near the point where $\gamma = \Gamma_O$ (optimal). The training curves are based on data from the first training epoch, the test curves on data from 3 training epochs.

(Section 5.1); or by an on-line procedure that takes $\gamma = \Gamma_P$ to be the point where $F(\gamma)+H(\gamma) = 1$ (Section 5.2). With a miss rate defined as $M(\gamma) \equiv 1 - H(\gamma)$, $F(\gamma) + H(\gamma) = 1$ at the point where the miss rate curve, which increases with γ, intersects the false alarm rate curve, which decreases with γ (Figure 2c). Thus for on-line simulations, the predicted miss rate $M_P \equiv M(\Gamma_P)$ equals the predicted hit rate $H_P \equiv H(\Gamma_P)$.

A new ROC curve $(F(\gamma), H(\gamma))$ is calculated for the test set data. Test set discrimination is performed using the threshold Γ_P calculated during training. The actual hit rate during testing is then $H_A \equiv H(\Gamma_P)$; the actual false alarm rate is $F_A \equiv F(\Gamma_P)$; and the actual miss rate is $M_A \equiv 1 - H_A$ (Figure 2c).

Finally, the actual hit and false alarm rates can be compared with optimal values H_O and F_O, which are obtained from *a posteriori* calculation of an optimal threshold Γ_O for the test set data. The optimal $(F(\Gamma_O), H(\Gamma_O))$ is, by definition, a point on the test set ROC curve that minimizes $[F(\gamma) - H(\gamma)]$. This point is typically close to the point where $F(\gamma) - H(\gamma) = 1$, where the false alarm rate curve intersects the miss rate false alarm curve (Figure 2c).

5.1 Validation set methods

One way to determine a predicted threshold Γ_P is by a procedure that partitions the training data into a training subset and a validation subset [11]. The network is trained on the training subset; then the ROC curve $(F(\gamma), H(\gamma))$ calculated for the validation subset takes Γ_P to be a point on the curve that minimizes $[F(\gamma) - H(\gamma)]$. For a familiarity discrimination task the validation set must include examples of classes not present in the training set. The value of the predicted threshold Γ_P is then employed for familiarity discrimination on the test set.

Table 1 shows results of this validation-without-retraining method applied to simulated range profile data. Predicted values of the hit and false alarm rates (H_P, F_P) are close to the actual rates (H_A, F_A), and these actual rates are in turn close to the optimal rates (H_O, F_O). Note that a network ideally should be trained on *all* available data before being used in the field. So, once Γ_P is determined, the training subset and validation subset should be recombined and the network retrained on the complete training set.

5.2 On-line threshold determination

The fuzzy ARTMAP network is capable of continuous learning during real-time operation, without a separation into training and testing phases. Although ARTMAP-FD with a predetermined threshold can also operate in a real-time mode, the validation-set method for deter-

	3x3			6x6			6x6*		
	predicted	actual	optimal	predicted	actual	optimal	predicted	actual	optimal
H	0.89	0.88	0.86	0.84	0.84	0.78	0.98	0.98	0.99
F	0.14	0.19	0.17	0.35	0.33	0.26	0.03	0.03	0.04
accuracy		0.94	1.00	–	0.92	1.00	–	1.00	1.00

Table 1: Predicted, actual, and optimal hit and false alarm rates, using threshold prediction by the validation-without-retraining method. Data sets are radar range profile simulations using 40 center frequencies for $wp \times wl$ simulated targets, where wp = number of wing positions and wl = number of wing lengths. The set of results for the 6x6* data set involves sequential evidence accumulation, with 100 observations per test target. Accuracy equals the fraction of correctly-classified targets out of familiar targets selected by the network as familiar. Training on half the target classes, validation on another quarter of the target classes, testing on all classes not present in the validation set.

mining the threshold is *not* compatible with real-time operation. Here we illustrate a method that determines Γ_P during a single epoch of training and that updates the value on-line while the network continues learning.

During ARTMAP-FD training, category nodes compete for new patterns as they are presented. When a node J wins the competition, learning expands the category hyperbox R_J enough to enclose the training pattern $\mathbf{a}$. The familiarity measure ϕ for each training set input then becomes equal to 1. However, *before* this learning takes place, ϕ can be less than 1, and the degree to which this initial value of ϕ is less than 1 reflects the distance from the training pattern to R_J. An event of this type—a training pattern successfully coded by a category node—is taken to be representative of familiar test-set patterns. The corresponding initial values of ϕ are thus used to generate a training hit rate curve, where $H(\gamma)$ equals the fraction of training inputs with $\phi > \gamma$.

What about false alarms? By definition, all patterns presented during training are familiar. However, a reset event during training (Section 2) resembles the arrival of an unfamiliar pattern during testing. Recall that a reset occurs when a category node that predicts class K wins the competition for a pattern that actually belongs to a different class k. The corresponding values of ϕ for these events can thus be used to generate a training false-alarm rate curve, where $F(\gamma)$ equals the fraction of match-tracking inputs with initial $\phi > \gamma$.

One way to improve predictive accuracy further is to use a reduced set of ϕ values in the training-set ROC curve construction process. Namely, training patterns that fall inside R_J, where $\phi = 1$, are not used because these exemplars tend to distort the miss rate curve. In addition, the *first* response to a training input is the best predictor of the network's response to a testing input, since sequential search will not be available during testing.

Finally, giving more weight to events occurring later in the training process improves accuracy. This can be accomplished by first computing training curves $M(\gamma)$ and $F(\gamma)$ and a preliminary predicted threshold Γ_P using the reduced training set; then recomputing the curves and Γ_P from data presented only after the system had activated the final category node of the training process (Figure 1c). The final predicted threshold Γ_P averages these values. This calculation can still be made on-line, by taking the "final" node to be the last one activated.

Table 2 shows that applying on-line threshold determination to simulated radar range profile data gives good predictions for the actual hit and false alarm rates, H_A and F_A. Most significantly, the threshold Γ_P predicted by this method gives H_A and F_A that are close to optimal, particularly when the ROC curve has a c-index close to one. The method is effective even when testing involves sequential evidence accumulation, despite the fact that the training curves use only single views of each target.

6 Discussion

ARTMAP-FD is seen to be capable of a high level of performance in both identification and familiarity discrimination in application to simulated multiwavelength radar range profiles of as many as 36 targets, especially when sequential evidence accumulation is employed. An on-line threshold prediction method can be used in place of off-line validation-set methods to

	3x3			6x6			6x6*		
	predicted	actual	optimal	predicted	actual	optimal	predicted	actual	optimal
H	0.75	0.81	0.86	0.79	0.77	0.77	0.79	0.99	0.98
F	0.25	0.11	0.14	0.21	0.24	0.23	0.21	0.06	0.02
accuracy		0.95	1.00	–	0.93	1.00	–	1.00	1.00

Table 2: Predicted, actual, and optimal hit and false alarm rates, using on-line threshold prediction. Data sets as in Table 1. Training on half the target classes, testing on all target classes.

determine the optimal value of the familiarity threshold.

The on-line method for threshold determination requires storage of the familiarity measure ϕ for each training pattern used in constructing the training miss rate and false alarm rate curves. In a dynamic environment in which the optimal threshold is changing, older training samples can be discounted with a weighting factor that decays with time. With many training patterns, on-line storage requirements for these calculations might be reduced by recording a histogram of ϕ values grouped into bins of γ ranges.

Acknowledgements

This research was supported in part by grants from the Advanced Research Projects Agency, the National Science Foundation, and the Office of Naval Research: NSF IRI-94-01659 (G. A. C.), ONR N00014-95-1-0657 (G. A. C.), ONR N00014-95-0409 (G. A. C., M. A. R.), and ONR N00014-96-0659 (M. A. R., W. W. S.). We would like to thank Marcos Campos for helpful comments.

Appendix: Radar range profiles

A radar range profile is a one-dimensional representation of a target, produced from a recording of a radar pulse reflection at high temporal resolution [5]-[7]. Several range profiles, constructed from the same view of the target but using pulses of different center frequencies, can be concatenated to form a multiwavelength radar range profile [8,9]. Simulations [1,9] use multiwavelength range profiles with center frequencies evenly spaced between 18GHz and 22GHz. The range-bin size is 2/3 m and the range profile covers 40m, so the number of components in a profile equals the number of center frequencies times 60.

References

[1] Carpenter, G.A., Rubin, M. A., & Streilein, W. W., ARTMAP-FD: Familiarity discrimination applied to radar target recognition, in *ICNN'97: Proceedings of the IEEE International Conference on Neural Networks*, Houston, June 1997.

[2] Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, D. B., Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps, *IEEE Transactions on Neural Networks*, **3**, 698-713, 1992.

[3] Carpenter, G. A., Grossberg, S., & Rosen, D. B., Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system, *Neural Networks*, **4**, 759-771, 1991.

[4] Carpenter, G. A., & Ross, W. D., ART-EMAP: A neural network architecture for object recognition by evidence accumulation, *IEEE Transactions on Neural Networks*, **6**, 805-818, 1995.

[5] Borden, B., Problems in airborne target recognition, *Inverse Problems*, **10**, 1009-1022, 1993.

[6] Hudson, S., & Psaltis, D., Correlation filters for aircraft identification from radar range profiles, *IEEE Transactions on Aerospace and Electronic Systems*, **19**, 741-748, 1993.

[7] Smith, C. R., & Goggans, P. M., Radar target identification, *IEEE Antennas and Propagation Magazine*, **35**, 27-38, 1992.

[8] Botha, E. C., Barnard, E., & Barnard, C. J., Feature-based classification of aerospace radar targets using neural networks, *Neural Networks*, **9**, 129-142, 1996.

[9] Rubin, M. A., Application of fuzzy ARTMAP and ART-EMAP to automatic target recognition using radar range profiles, *Neural Networks*, **8**, 1109-1116, 1995.

[10] Helstrom, C. W., *Elements of Signal Detection and Estimation*, Prentice Hall, Englewood Cliffs, NJ, 1995.

[11] Masters, T., *Practical Neural Network Recipes in C++*, Academic Press, San Diego, 1993.