ART-EMAP: A Neural Network Architecture for Learning and Prediction by Evidence Accumulation

Carpenter, Gail A.

Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems

http://hdl.handle.net/2144/1992

Boston University
ART-EMAP: A NEURAL NETWORK ARCHITECTURE FOR LEARNING AND PREDICTION BY EVIDENCE ACCUMULATION

Gail A. Carpenter and William D. Ross

January 1993


Permission to copy without fee all or part of this material is granted provided that: 1. the copies are not made or distributed for direct commercial advantage, 2. the report title, author, document number, and release date appear, and notice is given that copying is by permission of the BOSTON UNIVERSITY CENTER FOR ADAPTIVE SYSTEMS AND DEPARTMENT OF COGNITIVE AND NEURAL SYSTEMS. To copy otherwise, or to republish, requires a fee and/or special permission.

Copyright © 1993

Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems
111 Cummington Street
Boston, MA 02215
ART-EMAP: A Neural Network Architecture for Learning and Prediction by Evidence Accumulation

Gail A. Carpenter¹ and William D. Ross²
Center for Adaptive Systems and Department of Cognitive and Neural Systems
Boston University, 111 Cummington Street, Boston, Massachusetts 02215 USA

Abstract

This paper introduces ART-EMAP, a neural architecture that uses spatial and temporal evidence accumulation to extend the capabilities of fuzzy ARTMAP. ART-EMAP combines supervised and unsupervised learning and a medium-term memory process to accomplish stable pattern category recognition in a noisy input environment. The ART-EMAP system features (i) distributed pattern registration at a view category field; (ii) a decision criterion for mapping between view and object categories which can delay categorization of ambiguous objects and trigger an evidence accumulation process when faced with a low confidence prediction; (iii) a process that accumulates evidence at a medium-term memory (MTM) field; and (iv) an unsupervised learning algorithm to fine-tune performance after a limited initial period of supervised network training. ART-EMAP dynamics are illustrated with a benchmark simulation example. Applications include 3-D object recognition from a series of ambiguous 2-D views.

ART-EMAP: An ARTMAP System for 3-D Object Recognition

ART-EMAP (Fig. 1) is a neural network architecture that extends fuzzy ARTMAP (Carpenter, Grossberg, Markuzon, Reynolds, and Rosen, 1992) to accomplish target object or pattern class recognition in noisy or ambiguous input environments. During performance, ART-EMAP integrates spatial evidence distributed across coded recognition categories to predict a pattern class. When a decision criterion determines the pattern class choice to be ambiguous, additional input from the same unknown class is sought. Evidence from multiple inputs accumulates until the decision criterion is satisfied and a high confidence prediction can be made. Accumulated evidence can also be used by the predictive mapping to fine-tune the system during unsupervised rehearsal learning.

ART-EMAP was developed to address the problem of 3-D object recognition by 2-D view recognition. Applications would include a vision system capable of sampling different

¹Supported in part by British Petroleum (89-A-1204), DARPA (AFOSR 90-0083, and ONR N00014-82-J-4015), the National Science Foundation (NSF IRI 90-00530), and the Office of Naval Research (ONR N00014-91-J-4100).

²Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0083), British Petroleum (89-A-1204), the National Science Foundation (NSF IRI 90-00530), and the Office of Naval Research (ONR N00014-91-J-4100).
ART – EMAP Architecture

Figure 1: ART-EMAP architecture. The ARTMAP map field \( F^{ab} \) is replaced with a multi-field EMAP system. Distributed \( F^a_2 \) output pattern \( y^a \), resulting from partial contrast enhancement of \( F^a_2 \) input \( T^a \), is filtered through EMAP weights \( w^{a}_{jk} \) to determine the \( F^{ab}_1 \) activity \( x^{ab} \). If a predictive decision criterion (DC) is not met, additional input can be sought. Evidence from a sequence of input patterns belonging to the same unknown \( ART_b \) output class then accumulates at the EMAP field \( F^{ab}_2 \) until the decision criterion is met at \( F^{ab}_2 \).
2-D perspectives of 3-D objects. In this scenario, the scene generates an organized database of inputs which are sequential object views or noisy samples of the same view, spatially grouped in pattern classes. This approach to 3-D object recognition has been successfully used in neural network machine vision applications, particularly the aspect network (Baloch and Waxman, 1991; Seibert and Waxman, 1990). ART-EMAP further develops this strategy.

ART-EMAP systems are here illustrated on a DARPA benchmark simulation problem, circle-in-the-square (Wilensky, 1990). This problem requires a system to identify which points of a square lie inside and which lie outside a circle whose area is half that of the square. In the simulations, each new stage of the ART-EMAP architecture is evaluated using a single set of training/test exemplars. During a supervised learning phase, 100 randomly chosen circle-in-the-square points constitute a training set. System performance is evaluated on both noise-free and noisy test sets. The noise-free test set consists a discrete sampling of 11,000 points. The noisy test set is generated by adding random noise to each of the 11,000 inputs. The random noise is gaussian with standard deviation equal to 0.1 times the length of one side of the square.

Spatial Evidence Accumulation

ART-EMAP employs a spatial evidence accumulation process that integrates a distributed pattern of activity across coded category nodes to help disambiguate a noisy or novel input. Previous ART (Carpenter and Grossberg, 1987; Carpenter Grossberg, and Rosen, 1991) and ARTMAP (Carpenter, Grossberg, and Reynolds, 1991; Carpenter, Grossberg, Markuzon, Reynolds, and Rosen, 1992) simulations chose only the most highly activated category node at the field $F^a_i$ (Fig. 1) for recognition and prediction. During ART-EMAP performance, an input activates all $F^a_i$ nodes in proportion to a nonlinear measure of the match between the input pattern and each category's coded prototype vector. In the simulations, activity $y_j^a$ of the $j^{th}$ $F^a_i$ nodes obeys a normalized power rule:

$$y_j^a = \frac{(T_j^a)^p}{\sum_{n=1}^{N_a} (T_n^a)^p},$$

(1)

where $T_j^a$ is the net weighted input from $F^a_i$ to the $j^{th}$ $F^a_i$ node. The power rule (1) approximates the dynamics of a shunting competitive short-term-memory (STM) network that contrast-enhances its input pattern (Grossberg, 1973). The power rule is equivalent to the choice rule when $p$ is large. For smaller $p$, the distributed activity pattern uses information from the relative $F^a_i$ category activations to improve predictive performance at $ART_1$. The input $S_{k}^{ab}$ from $F^a_i$ to the $k^{th}$ $F^a_i$ node of the EMAP field $F_a^{ab}$ obeys the equation:

$$S_{k}^{ab} = \sum_{j=1}^{N_a} w_{jk}^{ab} y_j^a,$$

(2)

Since distributed activity at the $ART_a$ field $F^a_i$ (1) leads to distributed input to the EMAP field $F_a^{ab}$ (2), a means of choosing a winning prediction at the EMAP field $F_a^{ab}$ needs to be
Figure 2: Response plot decision boundaries, and performance accuracy for 100/11,000 training/test exemplars. Plotted points are those predicted to be outside the circle. Plots (a), (b), and (c) show decision boundaries in a noiseless test environment. Plots (d), (e), and (f) show performance with gaussian noise of $SD = 0.1$ added to each test input vector $a$. Plots (a) and (d) show fuzzy ARTMAP performance using the choice rule. Plots (b) and (e) show performance using the power rule with $p = 10$. Plots (c) and (f) show performance using the power rule with $p = 10$, a decision criterion $DC = 2.0$, and multiple views.
specified. The simplest method is to choose the EMAP category $K$ that receives maximal input from $F_2^a$. This can be implemented by letting $x_k^b = S_k^a$ and defining $F_2^b$ activity by:

$$y_K^b = \begin{cases} 1 & \text{if } x_K^b > x_k^b \text{ for all } k \neq K \\ 0 & \text{otherwise.} \end{cases}$$

(3)

Figure 2 shows how distributed $F_2^b$ activity improves performance on circle-in-the-square simulations. In the noise-free test environment (Fig. 2a,b), fuzzy ARTMAP, with choice at $F_2^a$, performs at 93.1% accuracy while ART-EMAP, with $p=10$ at $F_2^a$, performs at 95.7% accuracy. In the noisy test environment (Fig. 2d,e), fuzzy ARTMAP performs at 86.5% accuracy, while ART-EMAP performs at 88.4% accuracy.

**EMAP Predictive Decision Criterion**

A decision criterion (DC) imposed at the field $F_2^b$ can be used to delay $ART_b$ pattern class prediction when evidence for any one class is ambiguous. The decision criterion permits $ART_b$ prediction during unsupervised performance only when the most active EMAP category $K$ becomes a minimum proportion more active than the next most active category. Thus:

$$y_K^b = \begin{cases} 1 & \text{if } x_K^b > (DC) x_k^b \text{ for all } k \neq K \\ 0 & \text{otherwise,} \end{cases}$$

(4)

where $DC \geq 1$. When $DC > 1$, the decision criterion can prevent prediction by keeping all $y_k^b = 0$. This occurs when multiple EMAP categories are similarly activated at $F_1^b$, representing ambiguous predictive evidence. When no class $K$ is chosen by (4), additional inputs of the same (unknown) class are sought. In applications, additional inputs may come from noisy images at a fixed perspective (samples) or from ambiguous images at various perspectives (views). In circle-in-the-square simulations, views correspond to randomly chosen inputs from the same region, inside or outside the circle. Samples correspond to noise-altered versions of a single randomly chosen input. Using views as additional input with $DC=2.0$ during testing, ART-EMAP performance improves to 100% accuracy in the noiseless environment (Fig. 2c) and to 93.1% in the noisy environment (Fig. 2e).

**Temporal Evidence Accumulation**

Even using the decision criterion described above, each ART-EMAP prediction is ultimately based on evidence from just one input. Ambiguous inputs are simply discarded. Additional performance improvements can be achieved by integration of predictive evidence from a sequence of ambiguous inputs at the EMAP module. To realize this goal, a sequence of $F_1^b$ map activations is summed, accumulating evidence in the form of a medium-term memory (MTM) at the field $F_E^b$:

$$(T_k^b)^{\text{new}} = (T_k^b)^{\text{old}} + x_k^b$$

(Fig. 1). When the DC is met, $T_k^b$ is reset to 0. Activities $y_K^b$ at field $F_2^b$ obey:
Figure 3: Plots (a), (b), and (c) show ART-EMAP response for 100/11,000 training/test exemplars. Gaussian noise of $SD = 0.1$ was added to each test input vector $\mathbf{a}$. Plotted points were evaluated by the system to be outside the circle. In each simulation, $F^2$ activity is distributed, with $p = 10$. (a) Single view with $DC = 1.0$. (b) Multiple views. (c) Multiple samples. In (b) and (c), the DC decreases from 6.0 toward 1.0, by (7), as evidence accumulates, by the additive integration rule (5). During testing, $F^2$ activity is distributed, with $p = 10$ and DC = 1.0. Plots (d), (e) and (f) show decision boundaries in a noisless input environment. (d) No unsupervised rehearsal learning (as in Figure 2b) (e) Unsupervised rehearsal learning on 50 randomly chosen test set exemplars. (f) Unsupervised rehearsal learning on 900 randomly chosen test set exemplars.
\[ y^b_K = \begin{cases} 
1 & \text{if } T^b_K > (DC)T^a_k \\
0 & \text{otherwise.} 
\end{cases} \quad (6) \]

Temporal evidence accumulation (5)-(6) is equivalent to applying a decision criterion to a running average of map field activations \( x^b \) rather than to \( x^a \) itself, as in (4). Over multiple inputs, the influence of random noise is factored out. In (6), the DC starts large, but gradually decreases toward 1 to ensure that an \( ART_b \) class is eventually chosen.

In simulations, the DC decreased exponentially from 6.0 to 1.0:

\[ DC(l) = 5.0(1.0 - r)^{l-1} + 1.0, \quad (7) \]

where \( a^{(l)} \) is the \( l \)th input in a same-class sequence. The decay rate \( r \) was set equal to 0.2. Circle-in-the-square simulations show that ART-EMAP with evidence accumulation over multiple views improves test set performance from 88.4% (Fig. 3a) to 97.6% in the noisy environment. Evidence accumulation over multiple samples also improves performance, yielding 92.2% accuracy (Fig. 3c).

**Unsupervised Rehearsal Learning**

Temporal evidence accumulation allows the ART-EMAP system to recognize objects from a series of ambiguous views. However the system learns nothing from the final outcome of this decision process. If, for example, an input sequence \( a^{(1)}, \ldots, a^{(L)} \) predicts an \( ART_b \) category \( K \), by (5)-(6), the entire sequence would need to be presented again before the same prediction would be made.

Unsupervised rehearsal learning fine-tunes performance by feeding back to the system knowledge of the final prediction. Specifically, after input \( a^{(L)} \) allows ART-EMAP to choose the \( ART_b \) category \( K \), the sequence \( a^{(1)}, \ldots, a^{(L)} \) is re-presented, or rehearsed. Weights in an adaptive filter from \( F_2^b \) to \( F_2^b \) are then adjusted, shifting category decision boundaries so that each input \( a^{(l)} \) in the sequence becomes more likely, on its own, to predict category \( K \). Weights \( w^a_a, w^b_k, \) and \( w^a_k \), trained during the supervised learning interval, remain constant during unsupervised rehearsal learning.

Simulations illustrated in Figure 3 show how unsupervised rehearsal learning improves performance with noise-free inputs. Similar improvements occur with noisy inputs. Figure 3d shows ART-EMAP performance at 95.7% on single inputs, without unsupervised rehearsal learning, as in Figure 2b. In Figure 3e, rehearsal learning took place during presentation of 50 randomly chosen test inputs. Thereafter, predictive accuracy on single-input test set simulations increased to 96.2%. Rehearsal learning on a set of 900 inputs further improved performance to 96.6% (Fig. 3e). Small shifts in the decision boundary from Figure 3d to 3e and from Figure 3e to 3f can be seen.
References


