Artificial Neural Networks
T. Kohonen, K. Mäkisara, O. Simula and J. Kangas (Editors)

ARTMAP: A SELF-ORGANIZING NEURAL NETWORK ARCHITECTURE FOR FAST SUPERVISED LEARNING AND PATTERN RECOGNITION

Gail A. Carpenter, Stephen Grossberg, and John Reynolds†

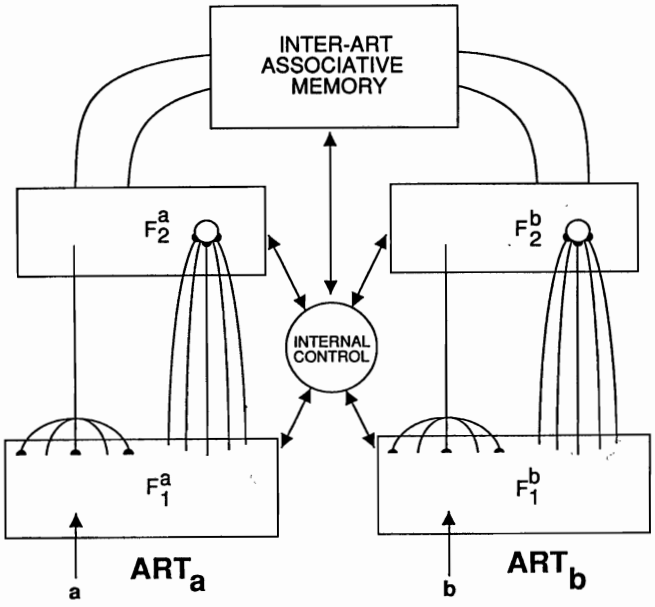
Center for Adaptive Systems
and
Department of Cognitive and Neural Systems
Boston University
111 Cummington Street
Boston, Massachusetts 02215 USA

1. Introduction

This paper announces a new neural network architecture, called ARTMAP [1], that autonomously learns to classify arbitrarily many, arbitrarily ordered vectors into recognition categories based on predictive success. This supervised learning system is built up from a pair of Adaptive Resonance Theory [2–5] modules (ART_a and ART_b) that are capable of self-organizing stable recognition categories in response to arbitrary sequences of input patterns (Figure 1). During training, the ART_a module receives a stream $\{\mathbf{a}^{(p)}\}$ of input patterns, and ART_b receives a stream $\{\mathbf{b}^{(p)}\}$ of input patterns, where $\mathbf{b}^{(p)}$ is the correct prediction given $\mathbf{a}^{(p)}$. These ART modules are linked by an associative learning network and an internal controller that ensures autonomous system operation in real time. During test trials, the remaining patterns $\mathbf{a}^{(p)}$ are presented without $\mathbf{b}^{(p)}$, and their predictions at ART_b are compared with $\mathbf{b}^{(p)}$.

Tested on a benchmark machine learning database in both on-line and off-line simulations, the ARTMAP system learns orders of magnitude more quickly, efficiently, and accurately than alternative algorithms, and achieves 100% accuracy after training on less than half the input patterns in the database. It achieves these properties by using an internal controller that conjointly maximizes predictive generalization and minimizes predictive error by linking predictive success to category size on a trial-by-trial basis, using only local operations. This computation increases the vigilance parameter ρ_a of ART_a by the minimal amount needed to correct a predictive error at ART_b. Parameter ρ_a calibrates the minimum confidence that ART_a must have in a category, or hypothesis, activated by an input $\mathbf{a}^{(p)}$ in order for ART_a to accept that category, rather than search for a better one through an automatically controlled process of hypothesis testing. Parameter ρ_a is compared with the degree of match between $\mathbf{a}^{(p)}$ and the top-down learned expectation, or prototype, that is read-out subsequent to activation of an ART_a category. Search occurs if the degree of match is less than ρ_a. ARTMAP is hereby a type of self-organizing expert system that calibrates the selectivity of its hypotheses based upon predictive success. As a result, rare but important events can be quickly and sharply distinguished even if they are similar to frequent events with different consequences. Between input trials ρ_a relaxes to a baseline vigilance $\overline{\rho_a}$. When $\overline{\rho_a}$ is large, the system runs in a conservative mode, wherein predictions are made only if the system is confident of the outcome. Very few false-alarm errors then occur at any stage of learning, yet the system reaches asymptote with no loss of speed. Because ARTMAP

† Acknowledgements: This research was supported in part by the Air Force Office of Scientific Research (AFOSR 90-0175 and AFOSR 90-0128), the Army Research Office (ARO DAAL-03-88-K0088), BP (98-A-1204), DARPA (AFOSR 90-0083), and the National Science Foundation (NSF IRI-90-00539). The authors wish to thank Cynthia E. Bradford for her valuable assistance in the preparation of the manuscript.



	Predictive ART	Back Propagation
supervised	yes	yes
self-organizing	yes	no
real-time	yes	no
self-stabilizing	yes	no
learning:	fast or slow match	slow mismatch

Figure 1. A Predictive ART, or ARTMAP, system includes two ART modules linked by an inter-ART associative memory. Internal control structures actively regulate learning and information flow. Back Propagation and Predictive ART both carry out supervised learning, but the two systems differ in many respects, as indicated.

learning is self-stabilizing, it can continue learning one or more databases, without degrading its corpus of memories, until its full memory capacity is utilized.

2. The ARTMAP System

The main elements of an ARTMAP system are shown in Figure 2. Two modules, ART_a and ART_b, read vector inputs **a** and **b**. If ART_a and ART_b were disconnected, each module would self-organize category groupings for the separate input sets. In the application described below, ART_a and ART_b are fast-learn ART 1 [2] modules coding binary input vectors. ART_a and ART_b are here connected by an inter-ART module that in many ways resembles ART 1. This inter-ART module includes a *Map Field* that controls the learning of

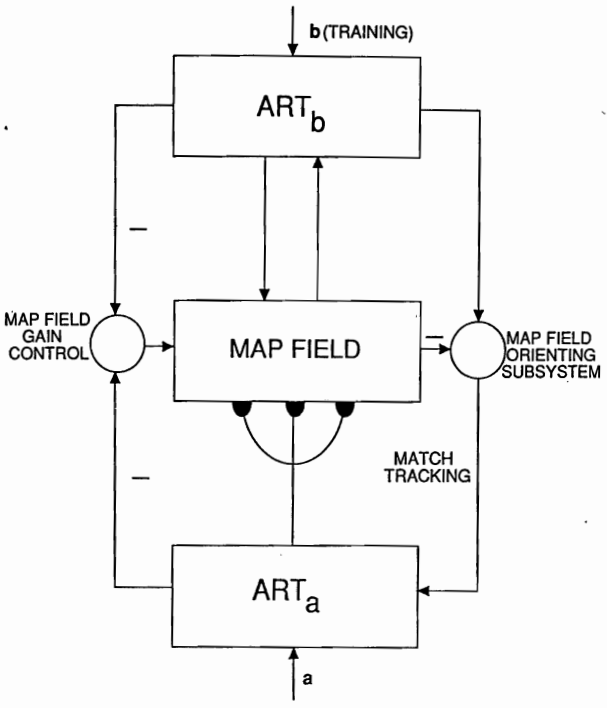


Figure 2. Block diagram of an ARTMAP system. Modules ART_a and ART_b self-organize categories for vector sets **a** and **b**. ART_a and ART_b are connected by an inter-ART module that consists of the Map Field and the control nodes called Map Field gain control and Map Field orienting subsystem. Inhibitory paths are denoted by a minus sign; other paths are excitatory.

an associative map from ART_a recognition categories to ART_b recognition categories. This map does not directly associate exemplars **a** and **b**, but rather associates the compressed and symbolic representations of families of exemplars **a** and **b**. The Map Field also controls match tracking of the ART_a vigilance parameter. A mismatch at the Map Field between the ART_a category activated by an input **a** and the ART_b category activated by the input **b** increases ART_a vigilance by the minimum amount needed for the system to search for and, if necessary, learn a new ART_a category whose prediction matches the ART_b category.

This inter-ART vigilance resetting signal is a form of "back propagation" of information, but one that differs from the back propagation that occurs in the Back Propagation network (Figure 1). For example, the search initiated by inter-ART reset can shift attention to a novel cluster of visual features that can be incorporated through learning into a new ART_a recognition category. This process is analogous to learning a category for "green bananas" based on "taste" feedback. However, these events do not "back propagate" taste features into the visual representation of the bananas, as can occur using the Back Propagation network. Rather, match tracking reorganizes the way in which visual features are grouped, attended, learned, and recognized for purposes of predicting an expected taste.

The following sections describe ARTMAP simulations using a machine learning benchmark database. For a full specification of the ARTMAP system, and analysis of network dynamics, see Carpenter, Grossberg, and Reynolds [1].

3. ARTMAP Simulations: Distinguishing Edible and Poisonous Mushrooms

The ARTMAP system was tested on a benchmark machine learning database that partitions a set of vectors **a** into two classes. Each vector **a** characterizes observable features of a mushroom as a binary vector, and each mushroom is classified as edible or poisonous [6]. The database represents the 11 species of genus *Agaricus* and the 12 species of the genus *Lepiota* described in **The Audubon Society Field Guide to North American Mushrooms** [7]. These two genera constitute most of the mushrooms described in the **Field Guide** from the familiy *Agaricaceae* (order *Agaricales*, class *Hymenomycetes*, subdivision *Basidiomycetes*, division *Eumycota*). All the mushrooms represented in the database are similar to one another: "These mushrooms are placed in a single family on the basis of a correlation of characteristics that include microscopic and and chemical features..." [7, p. 500]. The **Field Guide** warns that poisonous and edible species can be difficult to distinguish on the basis of their observable features. For example, the poisonous species *Agaricus californicus* is described as a "dead ringer" (p. 504) for the Meadow Mushroom, *Agaricus campestris*, that "may be known better and gathered more than any other wild mushroom in North America" (p. 505). This database thus provides a test of how ARTMAP and other machine learning systems distinguish rare but important events from frequently occurring collections of similar events that lead to different consequences.

The database of 8124 exemplars describes each of 22 observable features of a mushroom, along with its classification as poisonous (48.2%) or edible (51.8%). The 8124 "hypothetical examples" represent ranges of characteristics within each species; for example, both *Agaricus californicus* and *Agaricus campestris* are described as having a "white to brownish cap," so in the database each species has corresponding sets of exemplar vectors representing their range of cap colors. There are 126 different values of the 22 different observable features. For example, the observable feature of "cap-shape" has six possible values. Consequently, the vector inputs to ART_a are 126-element binary vectors, each vector having 22 1's and 104 0's, to denote the values of an exemplar's 22 observable features. The ART_b input vectors are (1,0) for poisonous exemplars and (0,1) for edible exemplars.

The ARTMAP system learned to classify test vectors rapidly and accurately, and system performance compares favorably with results of other machine learning algorithms applied to the same database. The STAGGER algorithm reached its maximum performance level of 95% accuracy after exposure to 1000 training inputs [8]. The HILLARY algorithm achieved similar results [9]. The ARTMAP system consistently achieved over 99% accuracy with 1000 exemplars, even counting "I don't know" responses as errors. Accuracy of 95% was usually achieved with on-line training on 300–400 exemplars and with off-line training on 100–200 exemplars. In this sense, ARTMAP was an order of magnitude more efficient than the alternative systems. In addition, with continued training, ARTMAP predictive accuracy always improved to 100%. These results are elaborated below.

Almost every ARTMAP simulation was completed in under 2 minutes on an IRIS 4D computer, with total time ranging from about 1 minute for small training sets to 2 minutes for large training sets. This is comparable to 2–5 minutes on a SUN 4 computer. Each timed simulation included a total of 8124 training and test samples, run on a time-sharing system with non-optimized code. Each 1–2 minute computation included data read-in and read-out, training, testing, and calculation of multiple simulation indices.

On-line learning simulations were carried out to imitate the conditions of a human or machine operating in a natural environment. An input **a** arrives, possibly leading to a prediction. If made, the prediction may or may not be confirmed. Learning ensues, depending on the accuracy of the prediction. Information about past inputs is available only through the present state of the system. Simulations of on-line learning by the ARTMAP system use each sample pair (**a**, **b**) as both a test item and a training item. Input **a** first makes a

TABLE 1: On-Line Learning

	Average number of correct predictions on previous 100 trials			
Trial	$\overline{\rho_a} = 0$ no replace	$\overline{\rho_a} = 0$ replace	$\overline{\rho_a} = 0.7$ no replace	$\overline{\rho_a} = 0.7$ replace
100	82.9	81.9	66.4	67.3
200	89.8	89.6	87.8	87.4
300	94.9	92.6	94.1	93.2
400	95.7	95.9	96.8	95.8
500	97.8	97.1	97.5	97.8
600	98.4	98.2	98.1	98.2
700	97.7	97.9	98.1	99.0
800	98.1	97.7	99.0	99.0
900	98.3	98.6	99.2	99.0
1000	98.9	98.5	99.4	99.0
1100	98.7	98.9	99.2	99.7
1200	99.6	99.1	99.5	99.5
1300	99.3	98.8	99.8	99.8
1400	99.7	99.4	99.5	99.8
1500	99.5	99.0	99.7	99.6
1600	99.4	99.6	99.7	99.8
1700	98.9	99.3	99.8	99.8
1800	99.5	99.2	99.8	99.9
1900	99.8	99.9	99.9	99.9
2000	99.8	99.8	99.8	99.8

Table 1: On-line learning and performance in forced choice ($\overline{\rho_a} = 0$) or conservative ($\overline{\rho_a} = 0.7$) cases, with replacement or no replacement of samples after training.

prediction that is compared with **b**. Learning follows as dictated by the internal rules of the ARTMAP architecture.

Four types of on-line simulations were carried out, using two different baseline settings of the ART_a vigilance parameter ρ_a: $\overline{\rho_a} = 0$ (forced choice condition) and $\overline{\rho_a} = 0.7$ (conservative condition); and using sample replacement or no sample replacement (Table 1). With sample replacement, any one of the 8124 input samples was selected at random for each input presentation. A given sample might thus be repeatedly encountered while others were still unused. With no sample replacement, a sample was removed from the input pool after it was first encountered. The replacement condition had the advantage that repeated encounters tended to boost predictive accuracy. The no-replacement condition had the advantage of having learned from a somewhat larger set of inputs at each point in the simulation. The replacement and no-replacement conditions had similar performance indices, all other things being equal. Each of the 4 conditions was run on 10 independent simulations. With $\overline{\rho_a} = 0$, the system made a prediction in response to every input. Setting $\overline{\rho_a} = 0.7$ increased the number of "I don't know" responses, increased the number of ART_a categories, and decreased the rate of incorrect predictions to nearly 0%, even early in training. The $\overline{\rho_a} = 0.7$ condition generally outperformed the $\overline{\rho_a} = 0$ condition, even when incorrect predictions and "I don't know" responses were both counted as errors. The primary exception occurred very early in training, when a conservative system gives the large majority of its no-prediction responses.

The data were tabulated to encode the number of correct predictions over the previous 100 trials (input presentations), averaged over 10 simulations. For example, with $\overline{\rho_a} = 0$

in the no-replacement condition, the system made, on the average, 94.9 correct predictions and 5.1 incorrect predictions on trials 201–300. In all cases a 95% correct-prediction rate was achieved before trial 400. With $\overline{\rho_a} = 0$, a consistent correct-prediction rate of over 99% was achieved by trial 1400, while with $\overline{\rho_a} = 0.7$ the 99% consistent correct-prediction rate was achieved earlier, by trial 800. Each simulation was continued for 8100 trials. In all four cases, the minimum correct-prediction rate always exceeeded 99.5% by trial 1800 and always exceeded 99.8% by trial 2800. In all cases, across the total of 40 simulations, 100% correct prediction was achieved on the last 1300 trials of each run.

A low correct-prediction rate for $\overline{\rho_a} = 0.7$ was made even on the first 100 trials. In the conservative mode, a large number of inputs initially make no prediction. With $\overline{\rho_a} = 0.7$ an average total of only 2 *incorrect* predictions were made on each run of 8100 trials. The asymptote of 100% accuracy was achieved faster than in the forced choice condition. Off-line learning simulations were equally successful.

4. Concluding Remarks

In summary, the ARTMAP system is designed to conjointly *maximize* generalization and *minimize* predictive error under *fast learning* conditions in *real time* in response to an *arbitrary ordering* of input patterns. Remarkably, the network can achieve 100% test set accuracy on a machine learning benchmark database, as described above. Each ARTMAP system learns to make accurate predictions quickly, in the sense of using relatively little computer time; efficiently, in the sense of using relatively few training trials; and flexibly, in the sense that its stable learning permits continuous new learning, on one or more databases, without eroding prior knowledge, until the full memory capacity of the network is exhausted.

References

[1] Carpenter, G.A., Grossberg, S., and Reynolds, J.H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, in press.

[2] Carpenter, G.A. and Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, **37**, 54–115.

[3] Carpenter, G.A. and Grossberg, S. (1987b). ART 2: Stable self-organization of pattern recognition codes for analog input patterns. *Applied Optics*, **26**, 4919–4930.

[4] Carpenter, G.A. and Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer*, **21**, 77–88.

[5] Carpenter, G.A., and Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, **3**, 129–152.

[6] Schlimmer, J.S. (1987a). Mushroom database. UCI Repository of Machine Learning Databases. (aha@ics.uci.edu)

[7] Lincoff, G.H. (1981). **The Audubon Society field guide to North American mushrooms**. New York: Alfred A. Knopf.

[8] Schlimmer, J.S. (1987b). Concept acquisition through representational adjustment (Technical Report 87–19). Doctoral dissertation, Department of Information and Computer Science, University of California at Irvine.

[9] Iba, W., Wogulis, J., and Langley, P. (1988). Trading off simplicity and coverage in incremental concept learning. In **Proceedings of the 5th international conference on machine learning**. Ann Arbor, MI: Morgan Kaufmann, 73–79.