ADAPTIVE RESONANCE THEORY

Gail A. Carpenter and Stephen Grossberg

Department of Cognitive and Neural Systems

Boston University

677 Beacon Street

Boston, Massachusetts 02215 USA

gail@bu.edu, steve@bu.edu

The Handbook of Brain Theory and Neural Networks, Second Edition
Michael A. Arbib, Editor
Cambridge, Massachusetts: MIT Press

Submitted: September, 1998
Revised: April, 2002

Short title (running head): Adaptive Resonance Theory

Correspondence:

Professor Gail A. Carpenter

Department of Cognitive and Neural Systems

Boston University

677 Beacon Street

Boston, Massachusetts 02215 USA

Phone: (617) 353-9483

Fax: (617) 353-7755

email: gail@bu.edu

INTRODUCTION

Principles derived from an analysis of experimental literatures in vision, speech, cortical development, and reinforcement learning, including attentional blocking and cognitive-emotional interactions, led to the introduction of adaptive resonance as a theory of human cognitive information processing (Grossberg, 1976). The theory has evolved as a series of real-time neural network models that perform unsupervised and supervised learning, pattern recognition, and prediction (Duda, Hart, and Stork, 2001; Levine, 2000). Models of unsupervised learning include ART 1 (Carpenter and Grossberg, 1987) for binary input patterns and fuzzy ART (Carpenter, Grossberg, and Rosen, 1991) for analog input patterns. ARTMAP models (Carpenter et al., 1992) combine two unsupervised modules to carry out supervised learning. Many variations of the basic supervised and unsupervised networks have since been adapted for technological applications and biological analyses.

MATCH-BASED LEARNING, ERROR-BASED LEARNING, AND STABLE FAST LEARNING

A central feature of all ART systems is a pattern matching process that compares an external input with the internal memory of an active code. ART matching leads either to a *resonant* state, which persists long enough to permit learning, or to a parallel memory search. If the search ends at an established code, the memory representation may either remain the same or incorporate new information from matched portions of the current input. If the search ends at a new code, the memory representation learns the current input. This *match-based learning* process is the foundation of ART code stability. Match-based learning allows memories to change only when input from the external world is close enough to internal expectations, or when something completely new occurs. This feature makes ART systems well suited to problems that require on-line learning of large and evolving databases.

Match-based learning is complementary to *error-based learning*, which responds to a mismatch by changing memories so as to reduce the difference between a target output and an actual output, rather than by searching for a better match. Error-based learning is naturally suited to problems such as adaptive control and the learning of sensory-motor maps, which require ongoing adaptation to present statistics. Neural networks that employ error-based learning include backpropagation and other multilayer perceptrons (MLPs) (Duda, Hart, and Stork, 2001; see BACKPROPAGATION).

Many ART applications use *fast learning*, whereby adaptive weights converge to equilibrium in response to each input pattern. Fast learning enables a system to adapt quickly to inputs that occur rarely but that may require immediate accurate recall. Remembering details of an exciting movie is a typical example of learning on one trial. Fast learning creates memories that depend upon the order of input presentation. Many ART applications exploit this feature to improve accuracy by voting across several trained networks, with voters providing a measure of *confidence* in each prediction.

CODING, MATCHING, AND EXPECTATION

Figure 1 illustrates a typical ART search cycle. To begin, an input pattern $\mathbf{I}$ registers itself as short-term memory activity pattern $\mathbf{x}$ across a field of nodes F_1 (Figure 1a). Converging and diverging pathways from F_1 to a coding field F_2, each weighted by an adaptive long-term memory trace, transform $\mathbf{x}$ into a net signal vector $\mathbf{T}$. Internal competitive dynamics at F_2 further transform $\mathbf{T}$, generating a compressed code $\mathbf{y}$, or *content-addressable memory*. With strong competition, activation is concentrated at the F_2 node that receives the maximal $F_1 \rightarrow F_2$ signal; in this *winner-take-all* (WTA) mode, only one code component remains positive (see WINNER-TAKE-ALL NETWORKS).

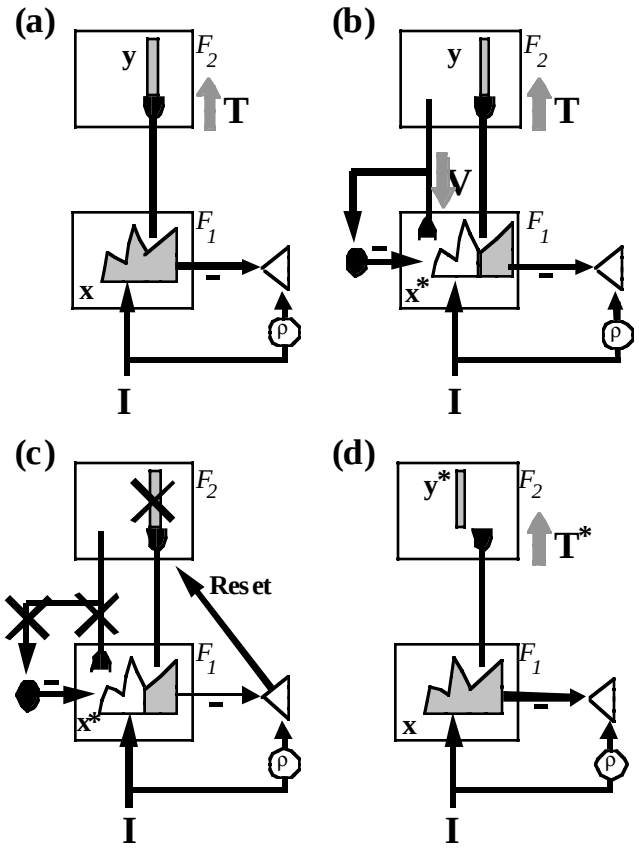

Figure 1. An ART search cycle imposes a matching criterion, defined by a dimensionless vigilance parameter ρ, on the degree of match between a bottom-up input **I** and the top-down expectation **V** previously learned by the F_2 code **y** chosen by **I**.

Before learning can change memories, ART treats the chosen code as a *hypothesis*, which it tests by matching the *top-down expectation* of **y** against the input that selected it (Figure 1b). Parallel specific and nonspecific feedback from F_2 implements matching as a real-time locally defined network computation. Nodes at F_1 receive both learned excitatory signals and unlearned

inhibitory signals from F_2. These complementary signals act to suppress those portions of the pattern $\mathbf{I}$ of bottom-up inputs that are not matched by the pattern $\mathbf{V}$ of top-down expectations. The residual activity $\mathbf{x}^*$ represents a pattern of *critical features* in the current input with respect to the chosen code $\mathbf{y}$. If $\mathbf{y}$ has never been active before, $\mathbf{x}^* = \mathbf{x} = \mathbf{I}$, and F_1 registers a perfect match.

ATTENTION, SEARCH, RESONANCE AND LEARNING

If the matched pattern $\mathbf{x}^*$ is close enough to the input $\mathbf{I}$, then the memory trace of the active F_2 code converges toward $\mathbf{x}^*$. The property of encoding an *attentional focus* of critical features is key to code stability. This learning strategy differentiates ART networks from MLPs, which typically encode the current input, rather than a matched pattern, and hence employ slow learning across many input trials to avoid catastrophic forgetting.

ART memory search begins when the network determines that the bottom-up input $\mathbf{I}$ is too novel, or unexpected, with respect to the active code to satisfy a matching criterion. The search process resets the F_2 code $\mathbf{y}$ before an erroneous association to $\mathbf{x}^*$ can form (Figure 1c). After reset, medium-term memory within the $F_1 \rightarrow F_2$ pathways (Carpenter and Grossberg, 1990) biases the network against the previously chosen node, so that a new code y^* may be chosen and tested (Figure 1d).

The ART matching criterion is determined by a parameter ρ called *vigilance*, specifies the minimum fraction of the input that must remain in the matched pattern in order for resonance to occur. Low vigilance allows broad generalization, coarse categories, and abstract memories. High vigilance leads to narrow generalization, fine categories, and detailed memories. At maximal vigilance, category learning reduces to exemplar learning. While vigilance is a free parameter in unsupervised ART networks, in supervised networks vigilance becomes an internally controlled variable which triggers search after rising in response to a predictive error. Because vigilance then varies across learning trials, the memories of a single ARTMAP system typically exhibit a range of

degrees of refinement. By varying vigilance, a single system can recognize both abstract categories, such as faces and dogs, and individual examples of these categories.

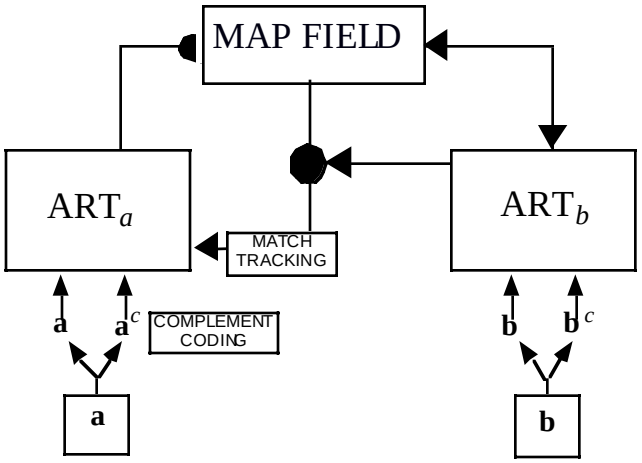

Figure 2. The general ARTMAP network for supervised learning includes two ART modules. For classification tasks, the ART_b module may be simplified.

SUPERVISED LEARNING AND PREDICTION

An ARTMAP system includes a pair of ART modules, ART_a and ART_b (Figure 2). During supervised learning, ART_a receives a stream of patterns $\left\{\mathbf{a}^{(n)}\right\}$ and ART_b receives a stream of patterns $\left\{\mathbf{b}^{(n)}\right\}$, where $\mathbf{b}^{(n)}$ is the correct prediction given $\mathbf{a}^{(n)}$. An associative learning network and a vigilance controller link these modules to make the ARTMAP system operate in real time, creating the minimal number of ART_a recognition categories, or *hidden units*, needed to meet accuracy criteria. A minimax learning rule enables ARTMAP to learn quickly, efficiently, and accurately as it conjointly minimizes predictive error and maximizes code compression in an on-line setting. A *baseline vigilance* parameter $\overline{\rho}_a$ sets the minimum matching criterion, with smaller $\overline{\rho}_a$ allowing broader categories to form. At the start of a training trial, $\rho_a = \overline{\rho}_a$. A predictive failure at ART_b increases ρ_a just enough to trigger a search, through a feedback control mechanism called *match tracking*. A newly active code focuses attention on a different cluster of input features, and

checks whether these features are better able to predict the correct outcome. Match tracking allows ARTMAP to learn a prediction for a rare event embedded in a cloud of similar frequent events that make a different prediction.

ARTMAP employs a preprocessing step called *complement coding*, which, by normalizing input patterns, solves a potential category proliferation problem (Carpenter, Grossberg, and Rosen, 1991). Complement coding doubles the number of input components, presenting to the network both the original feature vector and its complement. In neurobiological terms, complement coding uses both on-cells and off-cells to represent an input pattern. The corresponding on-cell portion of a weight vector encodes features that are consistently present in category exemplars, while the off-cell portion encodes features that are consistently absent. Small weights in complementary portions of a category representation encode as uninformative those features that are sometimes present and sometimes absent.

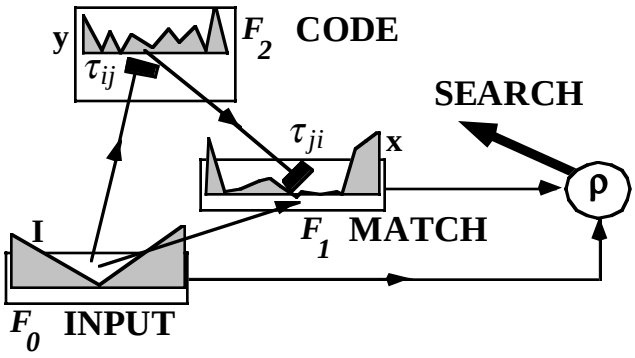

Figure 3. A distributed ART (dART) architecture retains the stability of WTA ART networks but allows the F_2 code to be distributed across arbitrarily many nodes.

DISTRIBUTED CODING

Winner-take-all activation in ART networks supports stable coding but causes category proliferation when noisy inputs are trained with fast learning. In contrast, distributed McCulloch-Pitts activation in MLPs promotes noise tolerance but causes catastrophic forgetting with fast learning (see LOCALIZED VS. DISTRIBUTED REPRESENTATIONS). *Distributed ART* (dART) models are designed to bridge these two worlds: distributed activation enhances noise tolerance while new system dynamics retain the stable learning capabilities of winner-take-all ART systems (Carpenter, 1997). These networks automatically apportion learned changes according to the degree of activation of each coding node, which permits fast as well as slow distributed learning without catastrophic forgetting.

New learning laws and rules of synaptic transmission in the reconfigured dART network (Figure 3) sidestep computational problems that occur when distributed coding is imposed on the architecture of a traditional ART network (Figure 1). The critical design element that allows dART to solve the catastrophic forgetting problem of fast distributed learning is the *dynamic weight*. This quantity equals the rectified difference between coding node activation and an *adaptive threshold*, thereby combining short-term and long-term memory in the network's fundamental computational unit.

Thresholds τ_{ij} in paths projecting directly from an input field F_0 to a coding field F_2 obey a *distributed instar* (dInstar) learning law, which reduces to an instar law when coding is WTA. Rather than adaptive gain, learning in the $F_0 \rightarrow F_2$ paths resembles the *redistribution of synaptic efficacy* (RSE) observed by Markram and Tsodyks (1996) at neocortical synapses. In these experiments, pairing enhances the strength, or efficacy, of synaptic transmission for low-frequency test inputs; but fails to enhance, and can even depress, synaptic efficacy for high-frequency test inputs. In the dART learning system, RSE is precisely the computational dynamic needed to support real-time stable distributed coding.

Thresholds τ_{ji} in paths projecting from the coding field F_2 to a matching field F_1 obey a distributed outstar (dOutstar) law, which realizes a principle of atrophy due to disuse to learn the

network's expectations with respect to the distributed coding field activation pattern. As in WTA ART systems, dART compares top-down expectation with the bottom-up input at the matching field, and quickly searches for a new code if the match fails to meet the vigilance criterion.

DISCUSSION: APPLICATIONS, RULES, AND BIOLOGICAL SUBSTRATES

ART and dART systems are part of a growing family of self-organizing network models that feature attentional feedback and stable code learning. Areas of technological application include industrial design and manufacturing, the control of mobile robots, face recognition, remote sensing land cover classification, target recognition, medical diagnosis, electrocardiogram analysis, signature verification, tool failure monitoring, chemical analysis, circuit design, protein/DNA analysis, 3-D visual object recognition, musical analysis, and seismic, sonar, and radar recognition (e.g., Caudell et al., 1994; Fay et al., 2001; Griffith and Todd, 1999). A book by Serrano-Gotarredona, Linares-Barranco, and Andreou (1998) discusses the implementation of ART systems as VLSI microchips. Applications exploit the ability of ART systems to learn to classify large databases in a stable fashion, to calibrate confidence in a classification, and to focus attention upon those featural groupings that the system deems to be important based upon experience. ART memories also translate to a transparent set of IF-THEN rules which characterize the decision-making process and which may be used for feature selection.

ART principles have further helped explain parametric behavioral and brain data in the areas of visual perception, object recognition, auditory source identification, variable-rate speech and word recognition, and adaptive sensory-motor control (e.g., Levine, 2000; Page, 2000). One area of recent progress concerns how the neocortex is organized into layers, clarifying how ART design principles are found in neocortical circuits (see LAMINAR CORTICAL ARCHITECTURE IN VISUAL PERCEPTION).

Pollen (1999) resolves various past and current views of cortical function by placing them in a framework he calls *adaptive resonance theories*. This unifying perspective postulates resonant feedback loops as the substrate of phenomenal experience. Adaptive resonance offers a core module for the representation of hypothesized processes underlying learning, attention, search, recognition, and prediction. At the model's field of coding neurons, the continuous stream of information pauses for a moment, holding a fixed activation pattern long enough for memories to change. Intrafield competitive loops fixing the moment are broken by active reset, which flexibly segments the flow of experience according to the demands of perception and environmental feedback. As Pollen (pp. 15-16) suggests: "it may be the consensus of neuronal activity across ascending and descending pathways linking multiple cortical areas that in anatomical sequence subserves phenomenal visual experience and object recognition and that may underlie the normal unity of conscious experience."

REFERENCES

Carpenter, G.A., 1997, Distributed learning, recognition, and prediction by ART and ARTMAP neural networks, Neural Networks, 10:1473-1494.

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. and Grossberg, S., 1990, ART 3: Hierarchical search using chemical transmitters in self–organizing pattern recognition architectures, Neural Networks, 3:129-152.

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 Transactions on Neural Networks, 3:698-713.

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

Caudell, T.P., Smith, S.D.G., Escobedo, R., and Anderson, M., 1994, NIRS: Large scale ART–1 neural architectures for engineering design retrieval, Neural Networks, 7:1339-1350.

* Duda, R.O., Hart, P.E., and Stork, D.G., 2001, Pattern Classification, Second Edition, New York: John Wiley, Section 10.11.2.

Fay, D.A., Verly, J.G., Braun, M.I., Frost, C., Racamato, J.P., and Waxman, A.M., 2001, Fusion of multi-sensor passive and active 3D imagery, in Proceedings of SPIE Vol. 4363 Enhanced and Synthetic Vision.

Griffith, N., and Todd, P.M. (Editors), 1999, Musical Networks: Parallel Distributed Perception and Performance, Cambridge, Massachusetts: MIT Press.

Grossberg, S., 1976, Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors & II: Feedback, expectation, olfaction, and illusions, Biological Cybernetics, 23:121-134 & 187-202.

* Levine, D.S., 2000, Introduction to Neural and Cognitive Modeling, Mahwah, New Jersey: Lawrence Erlbaum Associates, Chapter 6.

Markram, H., and Tsodyks, M., 1996, Redistribution of synaptic efficacy between neocortical pyramidal neurons, Nature, 382:807-810.

Page, M., 2000, Connectionist modelling in psychology: a localist manifesto, Behavioral and Brain Sciences, 23:443-512.

Pollen, D.A., 1999, On the neural correlates of visual perception, Cerebral Cortex, 9:4-19.

Serrano-Gotarredona, T., Linares-Barranco, B., and Andreou, A.G., 1998, Adaptive Resonance Theory Microchips: Circuit Design Techniques, Boston: Kluwer Academic Publishers.