ART Neural Networks: Distributed Coding and ARTMAP Applications

Gail A. Carpenter

Boston University, Department of Cognitive and Neural Systems, 677 Beacon Street, Boston, Massachusetts 02215 USA

gail@cns.bu.edu

Abstract. *ART (Adaptive Resonance Theory) neural networks for fast, stable learning and prediction have been applied in a variety of areas. Applications include airplane design and manufacturing, automatic target recognition, financial forecasting, machine tool monitoring, digital circuit design, chemical analysis, and robot vision. Supervised ART architectures, called ARTMAP systems, feature internal control mechanisms that create stable recognition categories of optimal size by maximizing code compression while minimizing predictive error in an on-line setting. Special-purpose requirements of various application domains have led to a number of ARTMAP variants, including fuzzy ARTMAP, ART-EMAP, Gaussian ARTMAP, and distributed ARTMAP. ARTMAP has been used for a variety of applications, including computer-assisted medical diagnosis. Medical databases present many of the challenges found in general information management settings where speed, efficiency, ease of use, and accuracy are at a premium. A direct goal of improved computer-assisted medicine is to help deliver quality emergency care in situations that may be less than ideal. Working with these problems has stimulated a number of ART architecture developments, including ARTMAP-IC [1]. This paper describes a recent collaborative effort, using a new cardiac care database for system development, has brought together medical statisticians and clinicians at the New England Medical Center with researchers developing expert systems and neural networks, in order to create a hybrid method for medical diagnosis. The paper also considers new neural network architectures, including distributed ART (dART), a real-time model of parallel distributed pattern learning that permits fast as well as slow adaptation, without catastrophic forgetting. Local synaptic computations in the dART model quantitatively match the paradoxical phenomenon of Markram-Tsodyks [2] redistribution of synaptic efficacy, as a consequence of global system hypotheses.*

Keywords. *Adaptive Resonance Theory, ART, ARTMAP, neural networks, medical prediction, redistribution of synaptic efficacy.*

1. ART and ARTMAP Neural Networks

Adaptive resonance theory originated from an analysis of human cognitive information processing and stable coding in a complex input environment [3,4]. An evolving series of ART neural network models have added new principles to the early theory and have realized these principles as quantitative systems that can be applied to problems of category learning, recognition, and prediction. Each ART network forms stable recognition categories in response to arbitrary input sequences with either fast or slow learning regimes. The first ART model, ART 1 [5], was an unsupervised learning system to categorize binary input patterns. ART 2 [6] and fuzzy ART [7] extend the ART 1 domain to categorize analog as well as binary input patterns.

Supervised ART architectures, called ARTMAP systems, self-organize arbitrary mappings from input vectors, representing features such as spectral values and terrain variables, to output vectors, representing predictions such as vegetation classes in a remote sensing application. Internal ARTMAP control mechanisms create stable recognition categories of optimal size by maximizing code compression while minimizing predictive error in an on-line setting. Binary ART 1 computations are the foundation of the first ARTMAP network [8], which therefore learns binary maps. When fuzzy ART replaces ART 1 in an ARTMAP system, the resulting fuzzy ARTMAP architecture [9] rapidly learns stable mappings between analog or binary input and output vectors.

2. Match-based Learning, Error-based Learning, and Fast Learning

The central feature of all ART systems is a pattern matching process that compares the current input with a learned expectation produced by an active code, or hypothesis. ART matching leads either to a resonant state, which focuses attention and triggers learning, or to a self–regulating parallel memory search, which eventually leads to a resonant state, unless the network's memory capacity is exceeded. If the search ends at an established code, the memory representation may stay the same or may be refined to incorporate information from attended portions of the current input. If the search ends at a new code, the code's memory representation begins by learning the current input itself. 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 and ARTMAP well suited to problems that require on–line learning of large and evolving databases.

Match–based learning is contrasted with *error–based learning*, which responds to a mismatch by sending the difference between a target output and an actual output toward zero, rather than by initiating a search 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 back propagation [10] and other multilayer perceptrons (MLPs).

Many ART applications use fast learning, whereby adaptive weights fully converge to equilibrium values in response to each input pattern. Fast learning enables a system to adapt quickly to inputs that occur only rarely but that may require immediate accurate recall. Remembering many details of an exciting movie is a typical example of fast learning. When the difference between actual output and target output defines "error," present inputs would drive out past learning, since fast learning zeroes the error on each input trial. Therefore fast learning destabilizes the memories of error–based models like back propagation. This feature restricts the domain of most MLPs to off–line applications with a slow learning rate.

3. Distributed Coding

In ART and ARTMAP networks, winner–take–all competitive activation supports stable coding, but this limiting case of competition may cause category proliferation when noisy inputs are trained with fast learning. In contrast, MLPs feature distributed McCulloch–Pitts activation, which promotes noise tolerance and code compression, but which causes catastrophic forgetting with fast learning. A recently introduced family of networks called *distributed ART* models combine the best of these two worlds: distributed activation enhances noise tolerance and code compression while new system dynamics retain the stable fast learning capabilities of winner–take–all (WTA) ART systems. With WTA coding, the unsupervised distributed ART model (dART) [11,12] reduces to fuzzy ART and the supervised distributed ARTMAP model (dARTMAP) [13] reduces to fuzzy ARTMAP. With distributed coding, these networks automatically apportion learned changes according to the degree of activation of each node, which permits fast as well as slow learning without catastrophic forgetting. A parallel distributed match–reset–search process also helps stabilize memory. The result of adaptation resembles long–term potentiation (LTP) for single-pulse or low-frequency test inputs but can resemble long-term depression (LTD) for higher frequencies. This dynamic is traced to dual computational properties of frequency-dependent and frequency-independent components of the coding signal. During learning, the frequency-independent component increases nonspecifically, for all inputs, while the frequency-dependent component becomes more selective, maximally favoring the current input.

The disappearance of LTP enhancement for high–frequency test inputs has been observed by Markram and Tsodyks [2] in the neocortex. Distributed ART features redistribution of synaptic efficacy at the local synaptic level as a consequence of necessary system hypotheses at the global pattern processing level. This "top-down" approach to understanding the Markram–Tsodyks data suggests, by example, how this apparently paradoxical phenomenon may actually be precisely the element needed to solve a critical pattern coding problem at a higher processing level. Both ART and dART models employ competitive learning schemes for code selection, and both are designed to stabilize learning. However, because ART networks use a classical steepest-descent paradigm called instar learning [14], these systems require winner-take-all coding to maintain stability with fast learning. A new learning law called the *distributed instar (dInstar)* allows dART code representations to be distributed across any number of network nodes. Both ART and dART also employ a preprocessing step called *complement coding* [7], which presents to the learning system both the original input vector and its complement. This device is analogous to on-cell/off-cell coding found in the early visual system. Complement coding solves a category proliferation problem pointed out by Moore [15]. It also suggests a computational solution to the tendency of redistribution of synaptic efficacy to enhance only low-frequency inputs: if an input component is consistently large with respect to a given code, then the network can embody this fact in the complementary component, which can also be enhanced since it will be consistently small.

The dynamic behavior of an individual dART synapse is seen in the context of its role in stabilizing distributed pattern learning, rather than as a primary hypothesis. Redistribution of synaptic efficacy here reflects a tradeoff between frequency-dependent and frequency-independent synaptic signal components which support a tradeoff between pattern selectivity and a nonspecific gain increase at the network level. Models that implement distributed coding via gain adaptation alone tend to suffer catastrophic forgetting and require slow or limited learning. In dART, each increase in frequency-independent synaptic efficacy is balanced by a corresponding decrease in frequency-dependent efficacy. The net result is redistribution, rather than nonspecific enhancement, of synaptic efficacy. The system uses this mechanism to enhance network response to a given pattern while suppressing the response to mismatched patterns. At the same time, the dART network learning law protects prior codes against catastrophic forgetting. It does so by formally replacing the traditional multiplicative weight with a *dynamic weight*, equal to the rectified difference between target node activation and an *adaptive threshold*, which embodies the long-term memory of the system [16]. The dynamic weight permits adaptation only at the most active coding nodes, which are limited in number due to competition at the target field. In addition, thresholds, which are initially zero, become increasingly resistant to change as they become larger. Note that, although thresholds following a minimal dInstar learning law can only increase monotonically, complement coding allocates two thresholds for each component of the original input, which allows the network to encode a full range of input features.

4. Rules, Applications, and Biological Substrates

ART principles have also been used to explain challenging behavioral and brain data in the areas of visual perception, visual object recognition, auditory source identification, variable–rate speech and word recognition, and adaptive sensory–motor control (e.g., [17,18]). One area of recent progress concerns how the neocortex is organized into layers. This new work suggests how "laminar computing" leads to intelligent behavior by modeling how bottom–up, top–down, and horizontal interactions are organized within the cortical layers. These interactions have thus far been studied within the visual cortex. Here, a model has been developed to show how visual cortex (1) stably develops circuits that match environmental constraints, and continues to refine this structure through adult learning; (2) binds or groups distributed information into coherent object representations; and (3) pays attention to important events (e.g., [19]). The mechanisms that govern (1) in the infant are proposed to lead to properties (2) and (3) in the adult. These results are clarifying how ART design principles are embedded within the neocortical circuits that subserve other types of intelligent behaviors, and open the way towards designing general–purpose vision systems that can autonomously learn optimal operating parameters in response to specialized image domains.

ART and dART systems are part of a rapidly growing family of attentive self–organizing systems that have evolved from the biological theory of cognitive information processing. ART modules have found their way into such diverse applications as 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., [20–22]). A recent book focuses on the implementation of ART systems as VLSI microchips [23].

Applications exploit the ability of an ART system to rapidly 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. The learned expertise of an ARTMAP system also translates to IF–THEN "rules." Within each recognition code, the expectation, or prototype represents a rule that predicts a given outcome. With WTA coding, these prototype vectors provide a transparent set of rules that characterize the decision–making process. ARTMAP neural networks have now provided new methodologies for medical database analysis. A case study of this method, applied to a cardiac database developed at the New England Medical Center (NEMC) [24] is introduced in the following section.

5. The New England Medical Center (NEMC) Modeling Project

A group of physicians and statisticians from the Division of Clinical Care Research at the New England Medical Center (NEMC) have created a database of Emergency Department patients who were considered for admission to the coronary care unit. A primary goal of the NEMC project is to develop methods to support a physician's decision making process. The project specifically aims to understand the utility and limitations of established and new modeling procedures and to promote their appropriate use in medical research, health care policy, and care assessment. These goals are accomplished through systematic investigation and rigorous evaluation of the relative predictive performance of the analyzed modeling methods. The project is being carried out as a collaboration between the physicians and statisticians who designed the NEMC cardiac database and researchers from Boston University and MIT.

5.1. The NEMC database

The NEMC cardiac database consists of the records of 3,068 study subjects examined at the Emergency Departments (ED) of six participating New England hospitals. The database includes clinical features available to ED physicians, such as clinical presentation, history, physical findings, electrocardiogram, socio-demographic characteristics, and coronary-disease risk factors. Of the 3,068 subjects in the database, 15.7% were diagnosed with cardiac problems requiring hospitalization. These positive outcome cases fall into three categories: arrhythmia, hemodynamic condition, and ischemia. These positive categories are not mutually exclusive; for example, arrhythmia and ischemia are usually accompanied by a hemodynamic condition. The identity of subcategories among the positive outcome variable was not used during model development. That is, the dichotomous output in the NEMC database codes only whether a patient required hospitalization, without specifying particular medical conditions.

The NEMC database includes records for each patient that represent 32 clinical variables, 199 raw ECG variables, and 78 derived ECG variables. Clinical variables quantify features such as medical complaints at arrival to the hospital, age, gender, body-mass index, history of past disease, and medication. The 199 raw ECG variables for the NEMC database were handpicked from a large pool of describing an ECG cycle. Chosen variables describe the amplitude, duration, slope, and area for segments of interest from each of the 12 leads (e.g., Q wave amplitude and duration, QRS area and duration, ST slope). Several variables describing general aspects of the ECG cycle were also included (e.g., mean ventricular rate, mean QRS duration, mean QT interval).

The 78 derived ECG variables in the NEMC database were created to separate clinically important aspects from irrelevant features of the raw signal. These variables are thought to be less sensitive to random fluctuations than the raw signal. Derived ECG variables consist of four groups:

- Five (5) derived ECG variables qualitatively describe the presence or absence of the following abnormalities in the cardiac cycle: Q waves, ST elevation, ST depression, T wave elevation, and T wave inversion. The derivation was based on recordings from all 12 leads.

- Fifty five (55) derived lead-by-lead ECG variables describe whether the five abnormalities enumerated above appear in individual leads. If an abnormality is detected in a raw lead, the same type of abnormality should be registered in one of its contiguous leads. Otherwise, the abnormality is attributed to noise and the variable is set to zero.

- Fifteen (15) summary regional ECG variables describe whether the locations of each of the five abnormalities, in the anterior, inferior, or lateral regions. For each region, at least two raw leads should have registered the abnormalities. Otherwise, the abnormality was ignored and the variable was set to zero.

- Three (3) summary dichotomous ECG variables describe the presence or absence of right bundle branch block, left bundle branch block, and left ventricular hypertrophy. According to the NEMC researchers, these variables have been used in previous regression models to override the effect of ST elevation, although it is not clear what their direct predictive value might be.

5.2. ARTMAP in the NEMC project

ARTMAP-IC [1], an extension of fuzzy ARTMAP [9] and ART-EMAP [25], was initially designed to solve a computational problems commonly encountered in medical modeling, including how to encode inconsistent cases, where identical patient records are associated with different outcomes. When the ARTMAP-IC system was introduced, its performance was evaluated on four benchmark medical databases. One of these databases was the Cleveland heart disease database from the UCI repository [26]. This database contained the records of 303 cardiology patients, 45.9% of which were diagnosed with heart disease. Each record had 13 attributes, including age, gender, heart rate, angina, ST depression, and ST slope. These initial simulation results demonstrated ARTMAP-IC's potential value for cardiac diagnosis.

Exploratory studies of the NEMC database indicated that ARTMAP-IC was not well suited for this problem. In particular, the low prevalence of positive outcomes (15.7%) rendered the system's instance counting feature counterproductive. The final successful ARTMAP algorithm for the NEMC project did, on the other hand, incorporate variations of the basic network that had been developed for other applications. This experience is typical: a given application usually benefits from certain model features but not others. A new model extension was also designed to improve the probability estimation capabilities of the ARTMAP system. Finally, although ARTMAP can handle an unlimited number of inputs, the significance of individual variables is clinically important. With 309 input variables in the NEMC database, variable selection and dimensionality reduction were important for purposes of interpretation. The project therefore included a novel method for estimation of the impact of individual input variables within the framework of ARTMAP networks. With this system, ARTMAP's generalization capabilities were seen to compare favorably to those of logistic regression and decision trees.

The logistic regression approach, on the other hand, appears to offer a simple model with reasonable discrimination and calibration capabilities. This claim of model simplicity is somewhat misleading because some of the derived variables, used as inputs to the regression, were laboriously handcrafted and have a great deal of complexity embedded in them. Still, one may argue that explicit variable derivation rules have certain advantages over complex self-organizing systems, such as ARTMAP or decision trees, because one can create rules encoding physicians' knowledge and diagnostic techniques. A counterargument is that self-organizing systems can discover unnoted patterns in the data and thus offer new diagnostic insights. A crucial test for broad acceptance of data-driven modeling approaches such as ARTMAP in the medical community would be the possibility of unraveling the information encoded in their complex structure. While an ARTMAP network is much easier to analyze and interpret than a standard backpropagation network, the high input dimensionality and the large number of category nodes pose ongoing challenges to structure visualization and rule extraction.

Acknowledgements: This research was supported in part by the Office of Naval Research (ONR N00014-95–1–0409 and ONR N00014–95–1–0657).

References

[1] Carpenter GA, Markuzon N. ARTMAP-IC and medical diagnosis: Instance counting and inconsistent cases. *Neural Networks* 1998; 11:323-336

[2] Markram H, Tsodyks M. Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature* 1996; 382:807–810

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

[4] Grossberg S. How does a brain build a cognitive code? *Psychological Review* 1980; 87:1-51

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

[6] Carpenter GA, Grossberg S. ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics* 1987b; 26:4919-4930

[7] Carpenter GA, Grossberg S, Rosen DB. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks* 1991; 4:759-771

[8] Carpenter GA, Grossberg S, Reynolds JH. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks* 1991; 4:565-588

[9] Carpenter GA, Grossberg S, Markuzon N, et al. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks* 1992; 3:698-713

[10] Werbos P. Beyond regression: New tools for prediction and analysis in the behavioral sciences. PhD thesis, Harvard University, Cambridge, Massachusetts, 1974

[11] Carpenter, GA. Distributed activation, search, and learning by ART and ARTMAP neural networks. In: *Proceedings of the International Conference on Neural Networks (ICNN'96): Plenary, Panel and Special Sessions*. IEEE Press, Piscataway, New Jersey, 1996, pp. 244-249

[12] Carpenter GA. Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. *Neural Networks* 1997; 10:1473-1494

[13] Carpenter, GA, Milenova, BL, Noeske, BW. Distributed ARTMAP: a neural network for fast distributed supervised learning. *Neural Networks* 1998; 11:793-813

[14] Grossberg, S. Neural expectation.: Cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes. *Kybernetik*, 1972; 10:49-57

[15] Moore, B. (1989). ART 1 and pattern clustering. In: Touretzky D, Hinton G, Sejnowski T (eds) *Proceedings of the 1988 Connectionist Models Summer School*. Morgan Kaufmann Publishers, San Mateo, California, 1989, pp. 174-185

[16] Carpenter, GA. A distributed outstar network for spatial pattern learning. *Neural Networks* 1994; 7:159-168

[17] Desimone R. Neural circuits for visual attention in the primate brain. In: Carpenter GA, Grossberg S (eds) *Neural Networks for Vision and Image Processing*. MIT Press, Cambridge, Massachusetts, 1992, pp 343-364

[18] Grossberg S. How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision* 1999a; 12:163-186

[19] Grossberg S. The link between brain learning, attention, and consciousness. *Consciousness and Cognition* 1999b; 8:1-44

[20] Caudell TP, Smith SDG, Escobedo R, et al. NIRS: Large scale ART–1 neural architectures for engineering design retrieval. *Neural Networks* 1994; 7:1339-1350

[21] Christodoulou CG, Huang J, Georgiopoulos M, et al. Design of gratings and frequency selective surfaces using fuzzy ARTMAP neural networks. *Journal of Electromagnetic Waves and Applications* 1995; 9:17-36

[22] Soliz P, Donohoe, GW. Adaptive resonance theory neural network for fundus image segmentation. In: *Proceedings of the World Congress on Neural Networks (WCNN'96)*, Erlbaum, Hillsdale, New Jersey, 1996, pp 1180-1183

[23] Serrano–Gotarredona T, Linares–Barranco B, Andreou, AG. *Adaptive Resonance Theory Microchips: Circuit Cesign Techniques*. Kluwer Academic Publishers, Boston, 1998

[24] Carpenter GA, Kopco N, Milenova BL, et al. A neural network method for supervised learning and medical database analysis. *Statistics in Medicine*, to appear

[25] Carpenter GA, Ross WD. ART-EMAP: A neural network architecture for object recognition by evidence accumulation. *IEEE Transactions on Neural Networks* 1995; 6:805-818

[26] Murphy PM, Aha DW. UCI repository of machine learning databases. University of California, Department of Information and Computer Science, Irvine, California. http://www.ics.uci.edu/~mlearn/MLRepository.html, 1992