1993-09-03

Adaptive Resonance Theory

ADAPTIVE RESONANCE THEORY

Gail A. Carpenter and Stephen Grossberg

September 1993

Technical Report CAS/CNS-94-034

To appear in:
Handbook of Brain Theory and Neural Networks
Arbib, M.A. (Ed.)
Cambridge, MA: MIT Press

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

ADAPTIVE RESONANCE THEORY

Gail A. Carpenter and Stephen Grossberg

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

September 23, 1993

Short title (running head): Adaptive Resonance Theory

Adaptive Resonance Theory, or ART, was introduced as a theory of human cognitive information processing (Grossberg, 1976a). The theory has since led to an evolving series of real-time neural network models whose unsupervised and supervised category learning, pattern recognition, and prediction remain stable in response to arbitrary input sequences with either fast or slow learning. Model families include the unsupervised models: ART 1 (Carpenter and Grossberg, 1987a) for learning to categorize binary input patterns; ART 2 (Carpenter and Grossberg, 1987b), ART2-A (Carpenter, Grossberg, and Rosen, 1991a), and fuzzy ART (Carpenter, Grossberg, and Rosen, 1991b) for learning to categorize either analog or binary input patterns; and ART 3 (Carpenter and Grossberg, 1990) for parallel search, or hypothesis testing, of distributed recognition codes in a multi-level network hierarchy. The supervised models include: ARTMAP (Carpenter, Grossberg, and Reynolds, 1991) for learning binary associative maps, and fuzzy ARTMAP (Carpenter *et al.*, 1992) for learning either analog or binary associative maps. These latter systems represent a computational synthesis of ideas from neural networks, expert production systems, and fuzzy logic. Model variations adapted to individual applications have been developed by a number of researchers.

Figure 1

Figure 1 illustrates one example from the family of ART 1 models, and Figure 2 illustrates a typical ART search cycle. Level F_1 in Figure 1 contains a network of nodes, each of which represents a particular combination of sensory features. Level F_2 contains a network of nodes that represent recognition codes which are selectively activated by patterns of activation across F_1. The activities of nodes in F_1 and F_2 are also called short term memory (STM) traces. STM is the type of memory that can be rapidly reset without leaving an enduring trace. For example, it is easy to reset the STM of a list of numbers that a person has heard once by distracting the person with an unexpected event. STM is distinct from LTM, or long term memory, which is the type of memory that we usually ascribe to learning. For example, we do not forget our parents' names when we are distracted by an unexpected event.

Figure 2

As shown in Figure 2a, an input vector **I** registers itself as a pattern **X** of activity across level F_1. The F_1 output vector **S** is then transmitted through the multiple converging and diverging adaptive filter pathways emanating from F_1. This transmission event multiplies the vector **S** by a matrix of adaptive weights, or LTM traces, to generate a net input vector **T** to level F_2. The internal competitive dynamics of F_2 contrast-enhance vector **T**. A compressed activity vector **Y** is thereby generated across F_2. In ART 1, the competition is tuned so that the F_2 node that receives the maximal $F_1 \rightarrow F_2$ input is selected. Only one component of **Y** is nonzero after this choice takes place. Activation of such a winner-take-all node defines the category, or symbol, of the input pattern **I**. Such a category represents all the inputs **I** that maximally activate the corresponding node. So far, these are the rules of a self-organizing feature map, also called competitive learning or learned vector quantization, as introduced in Grossberg (1972, 1976b) and von der Malsburg (1973).

Activation of an F_2 node may be interpreted as "making a hypothesis" about an input **I**. When **Y** is activated, it generates an output vector **U** that is sent top-down through the second adaptive filter. After multiplication by the adaptive weight matrix of the top-down filter, a net vector **V** inputs to F_1 (Figure 2b). Vector **V** plays the role of a learned top-down expectation. Activation of **V** by **Y** may be interpreted as "testing the hypothesis" **Y**, or "reading out the category prototype" **V**. The ART 1 network is designed to match the "expected prototype" **V** of the category against the active input pattern, or exemplar, **I**. Nodes that are activated by **I** are suppressed if they do not correspond to large LTM traces in the prototype pattern **V**. Thus F_1 features that are not "expected" by **V** are suppressed. This matching process changes the activity F_1 pattern **X** by suppressing activation of all the feature detectors in **I** that are not "confirmed" by hypothesis **Y**. The resultant pattern **X*** encodes the cluster of features in **I** that are relevant to hypothesis **Y** based upon the network's past experience. Pattern **X*** encodes the pattern of features to which the network "pays attention".

If expectation **V** is close enough to input **I**, then a state of *resonance* develops as the attentional focus takes hold. The pattern **X*** of attended features reactivates hypothesis **Y** which, in turn, reactivates **X***. The network locks into a resonant state through the mutual positive feedback that dynamically links **X*** with **Y**. The resonant state persists long enough for learning to occur; hence the term *adaptive resonance* theory. ART systems learn prototypes, rather than exemplars, because the attended feature vector **X***, rather than the input **I** itself, is learned.

This attentive matching process combines three different types of inputs at level F_1 (Figure 1): bottom-up inputs, top-down expectations, and attentional gain control signals. The attentional gain control channel sends the same signal to each F_1 node; it is a "nonspecific", or modulatory, channel. Attentive matching obeys a 2/3 Rule (Carpenter and Grossberg, 1987a): an F_1 node can be fully activated only if two of the three input sources that converge upon it send positive signals at a given time. The 2/3 Rule shows how an ART system can be "primed" by a previous event to expect a subsequent event that may or may not occur, and why priming is "unconscious". An active F_2 category and top-down expectation activates only to subthreshold levels the F_1 nodes in its prototype. No F_1 node can generate suprathreshold output signals in the absence of a "second third". Nodes are nonetheless "primed", or ready, to fire rapidly and resonantly if a bottom-up input does match the prototype well enough. Thus ART systems are "intentional" or "goal-oriented" systems that can selectively process expected events. Data from priming experiments during lexical decision and letter gap detection experiments have been explained using these 2/3 Rule operations (Grossberg, 1987).

The 2/3 Rule also allows an ART system to react to inputs in the absence of prior priming, since a bottom-up input both directly activates its target features and indirectly activates them via the nonspecific gain control channel to satisfy the 2/3 Rule (Figure 2a). After the input instates itself at F_1, leading to selection of a hypothesis $\mathbf{Y}$ and a top-down expectation $\mathbf{V}$, the 2/3 Rule ensures that only F_1 nodes that are confirmed by the expectation remain active in STM.

The criterion of an acceptable 2/3 Rule match is defined by a dimensionless parameter called *vigilance*. Vigilance weighs how close the input exemplar $\mathbf{I}$ must be to the top-down prototype $\mathbf{V}$ in order for resonance to occur. Because vigilance can vary across learning trials, recognition categories capable of encoding widely differing degrees of generalization, or morphological variability, can be learned by a single ART system. Low vigilance leads to broad generalization and abstract prototypes. High vigilance leads to narrow generalization and to prototypes that represent fewer input exemplars. At very high vigilance, prototype learning reduces to exemplar learning. Thus a single ART system may learn to recognize abstract categories of faces and dogs, as well as individual faces and dogs. In supervised ARTMAP systems, the prototypes that are learned depend upon the predictive success of their learned categories in a particular task environment.

If the top-down expectation **V** and the bottom-up input **I** are too novel, or unexpected, to satisfy the vigilance criterion, then a bout of hypothesis testing, or memory search, is triggered. Search leads to selection of a better recognition code, symbol, category, or hypothesis to represent input **I** at level F_2. An *orienting subsystem* mediates the search process (Figure 1). The orienting subsystem interacts with the attentional subsystem, as in Figures 2c and 2d, to enable the attentional subsystem to learn new F_2 representations with which to remember novel events without risking unselective forgetting of its previous knowledge.

The search process prevents associations from forming between **Y** and **X*** if **X*** is too different from **I** to satisfy the vigilance criterion. As shown in Figure 2c, the search process resets **Y** before such an association can form. A familiar category may be selected by the search if its prototype is similar enough to the input **I** to satisfy the vigilance criterion. The prototype may then be refined in light of new information carried by **I**. If **I** is too different from any of the previously learned prototypes, then an uncommitted F_2 node is selected and learning of a new category is initiated. A network parameter controls how deeply the search proceeds before an uncommitted F_2 node is chosen.

As inputs that correspond to a particular category are practiced over learning trials, the search process converges upon a stable learned recognition category in F_2. This process corresponds to making the inputs "familiar" to the network. After familiarization takes place, all inputs coded by that category access it directly in a one-pass fashion, and search is automatically disengaged. The category selected is, moreover, the one whose prototype provides the globally best match to the input pattern. Learning can proceed on-line, and in a stable fashion, with familiar inputs directly activating their categories, while novel inputs continue to trigger adaptive searches for better categories, until the network's memory capacity is fully utilized.

These properties of ART systems have been used to explain and predict a variety of cognitive and brain data that have, as yet, received no other theoretical explanation (Carpenter and Grossberg, 1991; Grossberg, 1987). For example, a formal lesion of the orienting subsystem creates a memory disturbance that mimics properties of medial temporal amnesia after lesions of the hippocampal formation (Carpenter and Grossberg, 1993). The F_1 and F_2 levels of the attentional subsystem are interpreted in terms of data concerning the prestriate visual cortex and the inferotemporal cortex during visual object recognition (Desimone, 1992), and the attentional gain control pathway is interpreted in terms of the pulvinar.

Figure 3

Supervised ARTMAP architectures (Carpenter, Grossberg, and Reynolds, 1991; Carpenter *et al.*, 1992) learn a map that associates a variable number of learned categories that compress one feature space (e.g., visual features) with learned categories of another feature space (e.g., auditory features). Figure 3 illustrates how ARTMAP can compress different sorts of information into distinct recognition categories which may make the same prediction. ARTMAP systems include a *match tracking* process that increases vigilance after a predictive failure just enough to trigger memory search for a better category. Multiple scales of generalization, from fine to coarse, are thereby created as needed. Match tracking realizes a Minimax Learning Rule that conjointly minimizes predictive error and maximizes generalization using only information that is locally available under incremental learning conditions in a nonstationary environment.

The expertise of an ARTMAP system can be inferred from the IF-THEN "rules" that it learns. Suppose, for example, that the input vectors encode biochemicals and that the output vectors encode drug effects on behavior. Various biochemicals may achieve the same clinical effect on behavior for different chemical reasons. Correspondingly, at any time during learning, the operator of an ARTMAP system can test how many recognition categories give rise to a desired clinical effect by checking which LTM traces are large in the pathways from input recognition categories to the desired output node. Within each recognition category, the prototype, or vector of large LTM traces, characterizes a particular "rule", or bundle of biochemical features, that predicts the desired clinical effect. The "IF-THEN" nature of the rule derives from the association between input and output categories; namely, "if the biochemical has features close enough to a particular prototype, then it predicts the desired outcome." A list of these prototype vectors provides a transparent set of rules that predict the desired outcome. Many such rules may coexist without mutual interference due to the competitive interactions whereby each hypothesis Y in Figure 2 is compressed. Associative networks such as back propagation often mix multiple rules among the same LTM traces because they do not have the competitive dynamics to separate them.

ARTMAP is one of a rapidly growing family of attentive self-organizing prediction systems that have evolved from the biological theory of cognitive information processing of which ART forms an important part (Carpenter and Grossberg, 1991). ART modules have found their way into such diverse applications as the control of mobile robots, a Macintosh system that adapts to user behavior, diagnostic monitoring systems for nuclear plants, learning and search of airplane part inventories, face recognition, target recognition, medical diagnosis, electrocardiogram analysis, protein/DNA analysis, 3-D visual object recognition, musical analysis, seismic recognition, sonar recognition, and laser radar recognition (e.g., Caudell *et al.*, 1991). All of these applications exploit the ability of ART systems to rapidly learn to classify large databases in a stable fashion, to calibrate their confidence in a classification, and to focus attention upon those featural groupings that they deem to be important based upon their past experience. The family of supervised ARTMAP systems is likely to find an even broader range of applications due to their ability to adapt the number, shape, and scale of their category boundaries to meet the on-line demands of large nonstationary databases.

Acknowledgments: The authors wish to thank Diana Meyers for her valuable assistance in the preparation of the manuscript. This research was supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225), ARPA (N00014-92-J-4015), the National Science Foundation (NSF IRI-90-00530), and the Office of Naval Research (ONR N00014-91-J-4100).

REFERENCES

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.

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.

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., and Grossberg, S. (Eds.), 1991, **Pattern Recognition by Self-Organizing Neural Networks**, Cambridge, MA: MIT Press.

Carpenter, G.A., and Grossberg, S., 1993, Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions, *Trends in Neurosciences*, 16:131--137.

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 Reynolds, J.H., 1991, ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network, *Neural Networks*, 4:565--588.

Carpenter, G.A., Grossberg, S., and Rosen, D.B., 1991a, ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, 4:493--504.

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

Caudell, T., Smith, S., Johnson, C., Wunsch, D., and Escobedo, R., 1991, An industrial application of neural networks to reusable design. **Adaptive Neural Systems**, Technical Report BCS-CS-ACS-91-001, Seattle, WA: The Boeing Company.

Desimone, R., 1992, Neural circuits for visual attention in the primate brain, in **Neural Networks for Vision and Image Processing**, (G.A. Carpenter and S. Grossberg, Eds.), Cambridge, MA: MIT Press, pp. 343-364.

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

Grossberg, S., 1976a, Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors, *Biological Cybernetics*, 23:121--134.

Grossberg, S., 1976b, Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions, *Biological Cybernetics*, 23:187--202.

Grossberg, S. (Ed.), 1987, **The Adaptive Brain, Volumes I and II**, Amsterdam: Elsevier/ North-Holland.

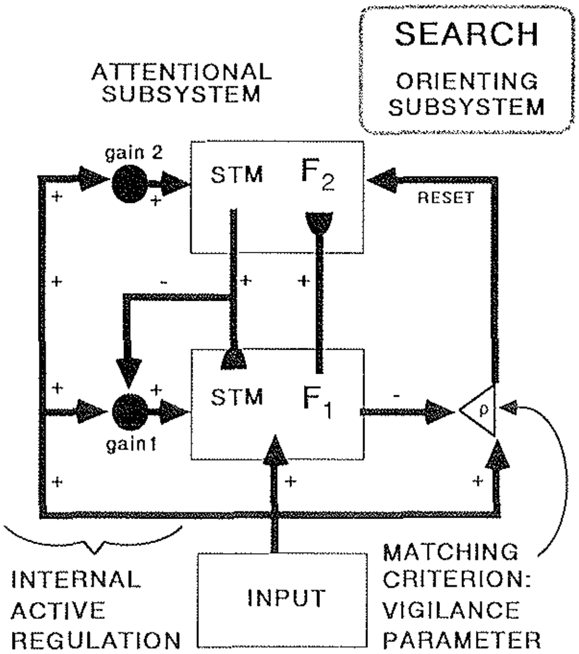
von der Malsburg, C., 1973, Self-organization of orientation sensitive cells in the striate cortex, *Kybernetik*, 14:85--100.

FIGURE CAPTIONS

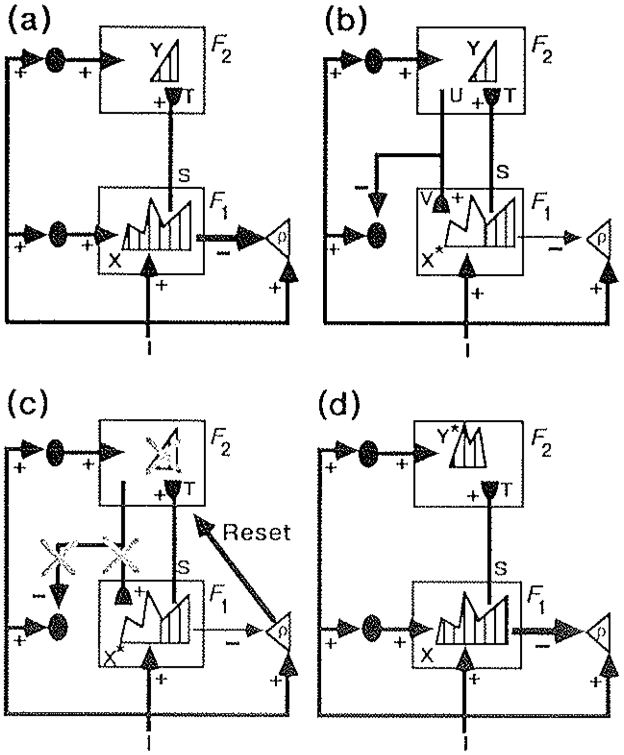
Figure 1. Typical ART 1 neural network (Carpenter and Grossberg, 1987a).

Figure 2. ART search cycle.

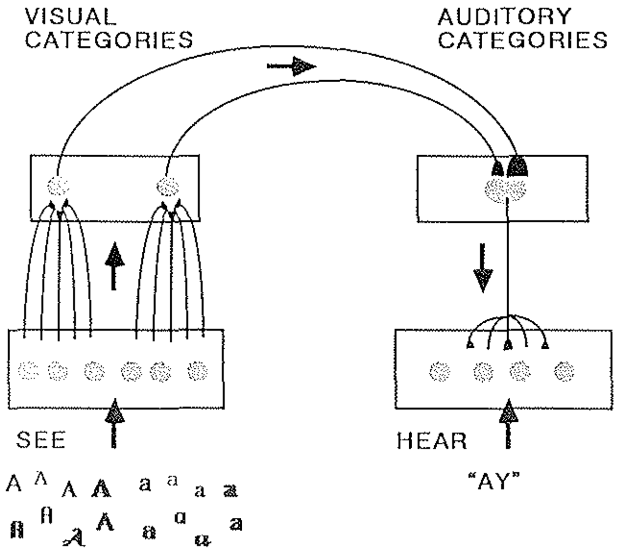
Figure 3. ARTMAP many-to-one learning combines categorization of many exemplars into one category, and labeling of many categories with the same name or prediction.



Carpenter and Grossberg Figure 1



Carpenter and Grossberg Figure 2



Carpenter and Grossberg Figure 3