Substance Use & Misuse, 32(2), 283–294, 1998

Adaptive Resonance Theory

Fabio Matera, Dr.

Semeion Research Center, Rome, Italy

INTRODUCTION

S. Grossberg (1976a, 1976b) and his partners at Boston's Center for Adaptive System have made an important contribution to the definition of neural networks models, notwithstanding the loss of credibility of the Connectionism, after the publication of the famous book "Perceptron" of M. Minsky and S. Papert (Minsky, 1969). This is particularly so, because their research concentrates upon the whole field of the brain functions, by developing numerous mathematical models of simulation.

"Recognition" is the identification process of the kind or category to which an object or an event belongs. The whole of all categories, with which we recognize the forms, constitutes a dynamic cognitive code that the brain automatically constructs as soon as we have the different experiences.

The Adaptive Resonance Theory (ART) (Carpenter, 1987a, 1988) is a model which tries to explain how this code is constructed, and how the brain, codifying the new experiences, utilizes the already existing categories. In other words, the ART is a self-organizing neural model, the result of the long history of models in this field, which attempts to overcome some constraints typical of these models. Furthermore, ART is subjected to further constraints imposed by the fact that Grossberg has always searched, in the functioning of his models, for the greatest similarity with biologic systems.

THE MODEL CONSTRAINTS

Biological plausibility imposes some constraints to the mathematical model's configuration and functioning:

- the neurons activate according to a differential equation (the membrane equation) not based on linear equations which describe the permeability, conductivity and activity of a biological membrane;

- the localization principle according to which information, in order to be transmitted, needs a physical sublayer. For example, the state of a synapsis can be determined only by components with which it is physically in contact. It is evident that non-local processes do not exist in biological processes;
- the stability/plasticity dilemma: the network has to be stable for the categories already classified, but also plastic enough so that when a new input is presented to it, it is able to "remember" the past, but at the same time to remain sensitive to the possible future.
- furthermore, the category's learning process must be "not supervised".

Two considerations about the model configuration from the analysis of these constraints arise which need to be clarified:

- a feedback process must be provided, in order to make sure of the stability of the model (top-down attentive feedback or template);
- a system must be introduced which "informs" ART that the new input is really new.

ART as Pattern Recognition

The ART networks are models for the classification of input successions. Theoretically, the goal of these networks is that of classifying, in an appropriate way, those inputs which are presented to it initially. It is important to note that the learning takes place without the presence of a supervisor or instructor. In other words, an ART network has all the information pertinent to the problem universe. It faces and discovers alone the crucial invariance of the classification of each input on its own.

There are two architectures: ART1 (Carpenter, 1988), for the classification of binary inputs, and ART2 (Carpenter, 1987b), for analogic models. We will describe only the ART1 network components.

THE ART1 ARCHITECTURE

The structure of the ART1 network is constituted by 4 interconnected systems:

1. the coding layer of the inputs (it takes out the input's characteristics), called F1;
2. the attentive or feedback layer in order to assure system stability, called F2;
3. a control system which allows the ART network to distinguish between an input signal and a signal, due perhaps to a noise, coming from F2 (gain control);
4. a system which allows the network to distinguish between a completely new input or which leads back to one of the already existing categories, called "novelty detector".

The heart of the structure is constituted by the F1 and the F2 layers which directly categorize the input successions, while the other two systems are useful in order to assure the system's stability because it works in the absence of a supervisor, and to prevent the same system from categorizing the noise in input.

The F1 Layer

As we said the heart of ART is formed by F1 and F2, which are two neuron layers connected for maximum gradient; each node of a layer is connected with all the nodes of the other. This doesn't mean that the two layers must necessarily have the same number of nodes, because, as we will see, each F2 node represents a category which codifies a succession of inputs.

Visually, image layers are arranged one upon the other with F1 downwards to the input connected and upwards to F2, while F2 is downwards with F1 connected (see Figure 1).

This structure predicts the presence of two weight matrices that we will call "bottom-up" (from F1 to F2) and "top-down" (from F2 to F1). Then, the F1 layer is connected to the input, by codifying its characteristics according to the strategy chosen by the operator. In fact, it can be decided that F1 simply represents the input, or is a manipulated version of the input itself through appropriate filters, allowing for each node of F1 to be activated only from particular zones of the input (there is an extensive bibliography in the literature about the filtering methodologies). In this second way F1 is assigned a preprocessing function.

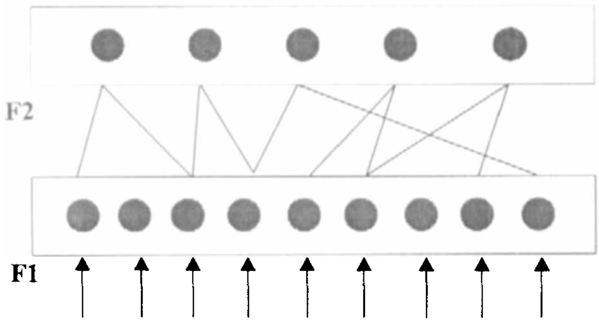

Figure 1.

In the first case F1 definitively codifies the input directly. In the second case F1 detects the fundamental characteristics (for the kind of filtering) of the input. Operatively, the input activates the F1 layer whose neurons are defined by Grossberg as "short term memory" (Grossberg, 1973), because the activation level of each neuron decreases very rapidly, in the absence of inputs, towards zero. This, according to the following equation, is a redefinition of the equation that describes the activity of biological membranes (Hodgkin, 1952):

$$(1) \quad \varepsilon \cdot \frac{dx_i}{dt} = -x_i + \left(1 - \mathrm{A} \cdot x_i\right) \cdot \Gamma_i^{+} - \left(1 - \mathrm{B} \cdot x_i\right) \cdot \Gamma_i^{-}$$

where: ε is a very small constant; x_i, activation value; (1-A $\cdot$ x_i) stops the activation at +1; (1-B $\cdot$ x_i) stops the activation at -1.

where:

$$(2) \quad \Gamma_i^{+} = \sum_{j \varepsilon G_E} w_{ij} f(x_j) + I_i$$

$$(3) \quad \Gamma_i^{-} = \sum_{j \varepsilon G_I} w_{ij} f(x_j)$$

The $f(x)$ function, is a patterning and nonlinear compression function with saturation. Saturating the activation value of a neuron means to prevent the F1 neuron activity from getting over the +1 and the -1 values. Positive saturation is determined by the second term, while the negative one is determined by the third term of formula (1). Equations (2) and (3) define the sets, Γ_i^+ and Γ_i^-, representing the contributions, as excitatory and inhibitory that the G_g (for the excitatory one) and the G_i (for the inhibitory one) sets to the activation of the i-th neuron of F1. This activation declines in time, according the (1). The w_{ij} terms represent the strength of the connections between the neurons of the two connected layers. The strength of these connections obeys the following differential equation:

$$(4) \qquad \frac{dw_{ij}}{dt} = K \cdot f(x_j) \cdot \left[-E_{ij} \cdot w_{ij} + h(x_i)\right]$$

where h is a patterning function (different from f), K is a constant, E_{ij} represents the coefficient, according to which the weight's activation, in the absence of input, declines. The functions choice of equations (2), (3) and (4), determines the task of F1: the direct coding of the inputs or the functioning as preprocessing.

The F2 Layer

Each neuron of the F1 layer activates, through the "bottom-up" connection matrix, the neurons of F2, according to the classic law of a neuron's activation (the sum weighed by the infralayer connections of the lower layer). Furthermore, as to the F2 layer which is assigned the role to categorize the inputs, only one among its neurons has to activate itself, and then is provided a WTA (Winner Takes All) algorithm of a competitive kind, determined according the formula:

$$(5) \qquad T_k = \sum w_{ki} \cdot x_{ki} = \sum_{i \varepsilon x} w_{ki}$$

where T_k represents the activation value of F2 k-th neuron; w_{ij} represents the connection between the F1 layer (i-th neuron); and the F2 layer (k-th neuron) represents x_{ki}. The formula reduces itself to the weights summation, because the activation values of F1 neurons will be equal to one if they belong to the X set which defines the input. Otherwise, they are null. Then, if the i-th neuron of the F2 layer finds the following condition:

(6) $T_i = \max_{k \varepsilon F2}\left\{T_k\right\}$

then it will be activated, while the others are put at zero. Also the neurons of F2 are defined as "short term memory" and then the value of their activation obeys the (1).

The activation of the winning node, through the top-down weights, strengthens or weakens the activation of F1 neurons that now and only now represents the network output. In fact, it is exactly from the interaction between the two layers (F1 and F2), in determining the output, that this theory has the name of Adaptive Resonance Theory. Furthermore, in order to solve the problem of stability, a trial is necessary to carry the F2 signal which makes up the output of F1 and then creates the resonance among the layers.

When the network is trained, each neuron of F2 represents a category. If, at this point, we present to the network a new entrance, it could happen that, in order to categorize the new input, the F2 layer will be completely modified. If a neuron changes the category which it represents time after time, this means that the network is not stable and therefore is not readable or credible.

But what does it mean that F2 represents a category? It means that for each input of a category, the only active neuron of F2 activates all the F1 neurons, characteristic of that input through the up-bottom connections, detecting in this way the invariant features of all inputs belonging to that category.

The "Gain Control" Subsystem

This system is useful in preventing the activation of an F2 neuron, probably due to the noise or to disturbances of other nature, which activates the F1 neurons. In other words, this system allows F1 to distinguish the signals coming from the up (disturbances) from those coming from the bottom (input succession). Otherwise, the activation of an F2 neuron would activate the F1 layer through the bottom-up connections. It, in its turn, would strengthen F2 through the bottom-up weights by creating the resonance. It is as if the network would have "hallucinations", an unacceptable condition in a network which has to categorize. The control system has three inputs and one output:

a. the input signal (positive);
b. the signal determined by the activation of F2 (negative);

c. an inhibition signal (negative);
d. an amplification signal of F1 activity (output of the system).

If there isn't any input, but if an F2 neuron is active, the signal which arrives to the control system will be negative or positive, but very small. Then, the amplification of the F1 activity is not enough to activate F2, and there isn't any redundancy phenomenon. In the case in which an input is present, the gain will be high, and then the F1 activity will allow the activation of F2.

The inhibition signal is a predetermined constant and, in the absence of input, unbalances the control system towards negative values by preventing a high value as amplifier of the F1 activity in exit. For example, when we are at a dinner, our attention is addressed more to our guest than to the food; this means that our system creates some conditions of attention priority. The same function can be assigned to the inhibition signal.

The "Novelty Detector" Subsystem

This subsystem is needed for preventing the network to learn and categorize inputs over the maximum possible number for that configuration (determined by the number of F2 neurons), primarily because the learning isn't supervised. The system of which we are speaking consists of two inputs and one output:

a. input to the network (positive);
b. activation value of F1 (negative);
c. "reset" signal of F2 neurons (output of the system).

When an input is presented to the network, it also enters in input to the novelty detector system; it activates the F1 layer and consequently the F2 layer, which gives back to F1 the resonance signal that modifies the activation of F1 neurons. This activation becomes the second input to the novelty detector. At this point, if the difference between the output signal and the input pattern goes over a one-dimensional threshold called "vigilance parameter", the novelty detector will send a reset signal to the neuron of the active F2 (it annuls it for the remaining learning) in order to prevent it from degrading the categorization before the new input, forcing the category to also include this input in its domain. Once the neuron resets, the process is repeated until either the input is categorized, or all neurons of F2 are reset. This means that the network has reached the maximum number of recognizable categories and it refuses the new input.

At this point, we can notice a limit of the ART network, because the introduction of the vigilance parameter, which has considerable effects on the categorization of the inputs, is externally decided. Then, it is clear that the definition of the parameter depends on the problem that one is facing.

THE LEARNING ALGORITHM

The learning phase of an ART network requires:

a. The activation of F1, on account of the presence of a non-null input. This activation could be the simple rewriting of the input or a filtered rewriting of it in relation to the configuration of the constructed network (preprocessing). This activation can also contribute the signal coming from an active F2 neuron for aleatory phenomena.
b. The activation of the F2 neurons according to the classic law of the neuron through the bottom-up weights. Successively, a process belonging to the WTA will activate a sole neuron by putting all the others at zero.
c. The updating of the bottom-up weights according to the F2 neuron which is winning.
d. The activation modification of F1 due to the signal sent by F2 through the top-down weights.
e. The updating of the top-down weights according to the input presented to the network or due to the consequent activation of F1.

In this work we have limited ourselves to only describing the ART1, which is a network with binary inputs (the extension to analogic inputs is fairly simple and it modifies neither the structure, nor the methodology of ART1), we indicate the input to the network with:

$$(7) \qquad I_i = \begin{cases} 1 & i \in Input \\ 0 & otherwise \end{cases}$$

named X the nodes set of F1

$$(8) \qquad X = \begin{cases} I & F2_{ij} = 0 \\ I \cap V^i & F2_{ij} \neq 0 \end{cases}$$

where V^i is the node set of the active F2.

The activation of the F2 nodes obeys the previous equations (5) and (6), for what concerns the reset rule of the active nodes of F2, and it is determined as:

$$\text{(9)} \qquad \frac{\left|I \cap V^i\right|}{|I|} < v$$

where v is the one dimensional vigilance parameter.

In order to complete the training rule, it is necessary to define the updating laws of the two weights matrices. These matrices are initialized as follows:

a. bottom-up, random between zero and the inverse of the model's number to learn;
b. top-down, $w_{ij} = 1$ for any i and j.

They are defined as "long term memory", because unlike what happens in the F1 and F2 neurons, their activity decreases very slowly in the absence of inputs according the differential equation (4). The updating equations for the bottom-up weights are:

$$\text{(10)} \qquad \frac{dw_{ji}}{dt} = \begin{cases} k \cdot \left[\left(1 - w_{ji}\right) \cdot L - w_{ji}\left(|x| - 1\right)\right] & i \in X \quad x_j \neq 0 \\ -k \cdot |x| \cdot w_{ji} & i \notin X \quad x_j \neq 0 \\ 0 & \quad\quad\;\; x_j = 0 \end{cases}$$

where the j pedex indicates the active neuron of F2, k is an amplification constant and L the decay constant of the weights value in absence of input.

The top-down weights follow the equation:

$$\text{(11)} \qquad \frac{dw_{ji}}{dt} = \begin{cases} -w_{ij} + & i \in X \quad x_j \neq 0 \\ -w_{ij} & i \notin X \quad x_j \neq 0 \\ 0 & \quad\quad\;\; x_j = 0 \end{cases}$$

It must be emphasized, that very often "fast" learning is utilized: it is supposed that equations (10) and (11) reach their asymptotic value much before a new pattern of input is presented to the network. In this way equations (10) and (11) are reduced for bottom-up weights:

$$(12)\quad w_{ji} = \begin{cases} \dfrac{L}{L-1+|x|} & i \in X \\ 0 & i \notin X \end{cases}$$

for the top-down weights:

$$(13)\quad w_{ij} = \begin{cases} 1 & i \in X \\ 0 & i \notin X \end{cases}$$

In both cases j represents the index of the winning F2 neuron. The updating equation of the bottom-up weights is very interesting in the case in which $i \in X$. Figure 2 shows this, in function to the L constant.

CONCLUSIONS

There are two strong points to the architecture of the ART model:

a. its unsupervised learning;
b. *the biological plausibility of its underpinning.*

In fact, the ART network is a model which tries to simulate the capability to categorize the external stimuli, emulating the behavior of living things by activating the neurons according to equations which are biologically "proper", and not using any kind of supervisor. In the ART network all of the input's world is represented by the "novelty detector" which has to decide if and into which category, among the ones at its disposal, to insert the new input.

This is also a major limitation of the network. This is so because all the above described operations depend on the vigilance parameter. This parameter is (once the value's chosen) completely extraneous to the ART architecture.

From this point of view, we can say that ART is not an adaptive system. In fact, if we choose as inputs an I and an O, ART will categorize them in the right way; on the other hand, if we make, in an appropriate way, the two letters noisy (in the same quantity), transforming the O into a Q, the introduced noise will weigh, in percentage, more on the I than on the O: or the contrary can happen.

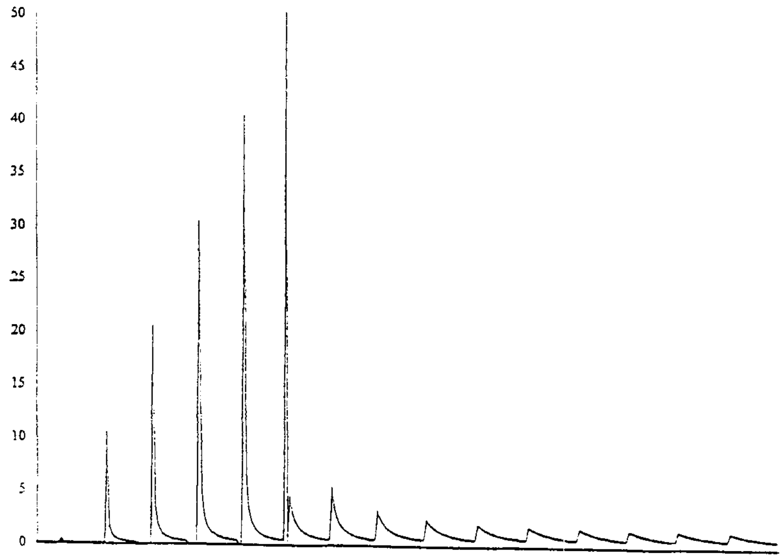

Figure 2.

Furthermore, if we compare the ART network with other classification networks, such as the Recirculation Networks (RC) with the Re-Entry technique created by Semeion Research Center of Rome (Buscema, 1994, 1995, 1996), the results of ART are very poor, both in terms of quality and resolution time of the problem.

Actually, the ART network (in different configurations) is utilized as part of complex neural networks architectures (Carpenter, 1989). Lastly, we note the possibility of combining the ART architectures, in an appropriate way, in order to achieve stable associative memories.

REFERENCES

BUSCEMA, 1994: M. Buscema, *Squashing Theory. Modello a Reti Neurali per la Previsione dei Sistemi Complessi*, Collana Semeion, Armando, Rome, 1994 [Squashing Theory: A Neural Network Model for Prediction of Complex Systems, Semeion Collection by Armando Publisher].

BUSCEMA, 1995: M. Buscema, *Constraint Satisfaction and Recirculation Neural Networks*, T.P. n.18, Semeion, Rome, 1995.

BUSCEMA, 1996: M. Buscema, *SQUASH 1.0.*, Shell for program Feed Forward Neural Networks, Semeion Software n. 5, Rome, 1992-1996.

CARPENTER, 1987a: G. A. Carpenter and S. Grossberg, *A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine*, Computer Vision, Graphics and Image Processing, vol. 37, pp. 54–115, 1987.

CARPENTER, 1987b: G. A. Carpenter and S. Grossberg, *ART2: Self-Organizing of Stable Category Recognition Codes for Analog Input Pattern*, Applied Optics, vol. 26, pp. 4919–4930, 1987.

CARPENTER, 1988: G. A. Carpenter and S. Grossberg, *The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network*, Computer, March 1988, pp. 77–88.

CARPENTER, 1989: G. A. Carpenter, S. Grossberg, and C. Mehanian, *Invariant Recognition Scenes by Self-Organizing ART Architecture: CORT-X Boundary Segmentation*, Neural Networks, vol. 2, pp. 169–181, 1989.

GROSSBERG, 1973: S. Grossberg, *Contour, Enhancement, Short Term Memory, and Costancies in Recognition Codes for Analog Input Pattern*, Studies in Applied Mathematics, vol. 52, pp. 217–257, 1973.

GROSSBERG, 1976a: S. Grossberg, *Adaptive Pattern Classification and Universal Recoding, I: Parallel Development and Coding of Neural Feature Detectors*, Biological Cybernetics, vol. 23, pp. 121–134, 1976.

GROSSBERG, 1976b: S. Grossberg, *Adaptive Pattern Classification and Universal Recoding, II: Feedback, Expectation, Olfaction and Illusion*, Biological Cybernetics, vol. 23, pp. 187–202, 1976.

HODGKIN, 1952: A. L Hodgkin and A. F. Huxley, *A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve*, Journal of Physiology, vol. 117, pp. 500–544, 1952.

KOHONEN, 1980: T. Kohonen, *Content Addressable Memories*, Springer Verlag, Berlin, 1980.

KOHONEN, 1984: T. Kohonen, *Self-Organization and Associative Memories*, Springer Verlag, Berlin, 1984.

MINSKY, 1969: M. Minsky and S. A. Papert, *Perceptrons*, The MIT Press, Cambridge, MA, (extended edition, 1988).

THE AUTHOR

Fabio Matera, Dr., degree in Electronic Engineering with par-ticular regard to the automatic and industrial aspects. His concentra-tion of study and experimental dissertation, realized with the E.N.E.A. (Federal Corporation for the New Technologies, Energy and Environment studies), were on Neural Networks as preprocessing filter for the segmentation of images. Researcher of the Semeion Research Center up to 1996.