Journal of Computer Science 3(11): 869-873, 2007
ISSN 1549-3636

Adaptive Resonance Theory 1 (ART1) Neural Network Based Horizontal and Vertical Classification of 0-9 Digits Recognition

Mbaitiga Zacharie
Department of Media Information Engineering, Okinawa National College of Technology,
905 Henoko, Nago, 905-2192, Okinawa Prefecture, Japan

Abstract: This study describes the Adaptive Resonance Theory 1 (ART1), an efficient algorithm that emulates the self-organizing pattern recognition and hypothesis testing properties of the ART neural network architecture for horizontal and vertical classification of 0-9 digits recognition. In our approach the ART1 model can self-organize in real time producing stable and clear recognition while getting input patterns beyond those originally stored. It can also preserve its previously learned knowledge while keeping its ability to learn new input patterns that can be saved in such a fashion that the stored patterns cannot be destroyed or forgotten. A parameter called the attentional vigilance parameter determines how fine the categories will be. If vigilance increases or decreases due to environmental control feedback, then the system automatically searches for and learns fine recognition categories.

Key words: ART1, gain control, input pattern, signal

INTRODUCTION

A formal analysis of how to overcome the learning instability experienced by a competitive learning model let to the introduction of an expended theory, called adaptive resonance theory (ART). This formal analysis showed that a certain type of top-down learned feedback and matching mechanism could significantly overcome the instability problem. It was realized that top-down attentional mecha-nisms, which had earlier been discovered through an analysis of interactions between cognitive and reinforcement mechanisms (Grossberg, 1975), had the same properties as these code-stabilizing mechanisms. In other words, once it was recognized how to formally solve the instability problem, it also became clear that one did not need to invent any qualitatively new mechanisms to do so. One only needed to remember to include previously discovered attentional mechanisms. These additional mechanisms enable code learning to self-stabilize in response to an essentially arbitrary in put environment. The basic principles of the adaptive resonance theory were introduced by Grossberg[1]. A class of ART called ART1, has since characterized as a system of ordinary differential equations by Carpenter and Grossberg[2,3]. Theorems have been proved that trace the real-time dynamics of ART 1 networks in response to arbitrary sequences of binary input patterns. These theorems predict both the order of search, as a function of the learning history of the network and the asymptotic category structure self-organized by arbitrary in-put patterns. They also prove the self-stabilization property and showed that the system's adaptive weights oscillate at most once and do not get trapped in spurious memory states or local minima. The novelty of this study lies on the horizontal and vertical classification of 0-9 digits recognition using a fast learning ART1, where the ART1 can self-organize, recognize these digits horizontally and vertically at the same time within a short period of time. It can also change the order of these digits during the learning process according to the training method, where the vigilance parameter determines how fine the category will be. Gain control enables the architecture to suppress noise up to a prescribe level. The result of this work will be introduced soon into the vision sensor for the recognition of the road vehicle numbers plate.

ART1 model description: ART1 is an unsupervised learning model specially designed for recognizing binary patterns. It typically consists of an attentional subsystem, an orienting subsystem as shown in Fig. 1, a vigilance parameter and a reset module. The vigilance parameter has considerable influence on the system. High vigilance produces higher detailed memories such as fine categories etc, while lower vigilance results in more general memories. The ART1 attentional has two competitive networks, comparison field layer F1 and the recognition field layer F2, two control gains, Gain1 and Gain2 and two short-term memory (STM) stages F1 and F2. Long-term memory (LTM) traces between F1 and F2 multi-ply the signal in these pathways.

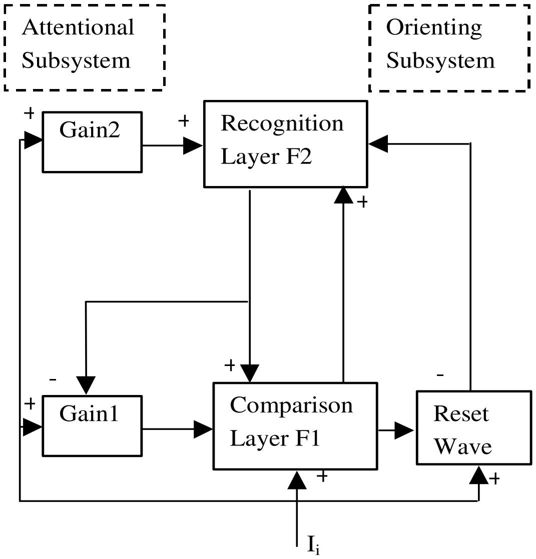

Fig. 1: Adaptive Resonance Theory Structure

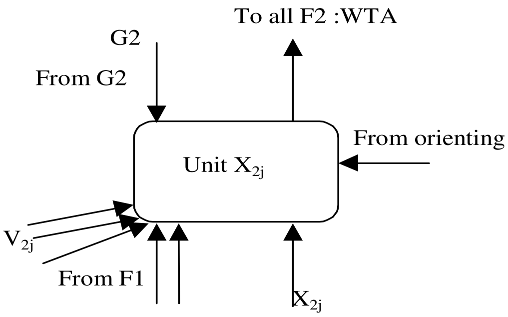

Fig. 2: Processing element X_{2j} in layer F2

Gains control enables F1 and F2 to distinguish current stages of running cycle. STM reset wave inhibits active F2 cells when mismatches between bottom-up and top-down signal occur at F1. The comparison layer receives the binary external input passing it to the recognition layer responsible for matching it to a classification category. This result is passed back to the comparison layer to find out if the category matches that of the input vector. If there is a match a new input vector is read and the cycle start again. If there is a mismatch the orienting system is in charge of inhibiting the previous category in order to get a new category match in the recognition layer. The two gains control the activity of the recognition and the comparison layer respectively. The orienting subsystem generates a reset wave to F2 when the bottom-up input pattern and top-down template pattern at F1, according to the vigilance criterion. The reset wave selectively and enduringly inhibits active F2 cell until the current is shut off. Offset of the input pattern terminates its processing at F1 and triggers offset of Gain2. Gain2 offset causes rapid decay of STM at F2 and thereby prepares F2 to encode the next input pattern without bias.

ART1 implementation: As state above, ART1 is a self organizing neural network having input and output neurons mutually coupled via bottom-up and top-down adaptive weights that perform recognition. To begin our approach, the network is first trained in accordance with the adaptive resonance theory by inputting reference pattern data under the form of 5×5 matrix (the very novelty of this study) into the neurons for clustering within the output neurons. Next, the maximum number of nodes in F2 is defined following by the vigilance parameter. The inputted pattern registered itself as short-term memory activity across a field of nodes F1. Converging and diverging pathways from F1 to a coding field F2, each weighted by an adaptive long-term memory trace, transform into a net signal vector T. Internal competitive dynamics at F2 further transform T, generating a compressed code or content addressable memory. With strong competition, activation is concentrated at the F2 node that receives the maximal F1→F2 signal. The main focus of this work is divided in four phases as follows: Comparison, recognition, search and learning.

Comparison: In the comparison or top-down template matching the short -time memory (STM) activation patternX_2 on F2 as shown in Fig. 2 generates a top-down template on F1. This STM is the reference input that we have inputted into the neurons during the implementation.

Then this pattern is multiplied by the long-term memory (LTM) traces V_{12} connecting from F2 to F1. Each node in-F1 sums up all its LTM gate signals.

$$V_{li} = \sum_j x_{2j} w_{21ij} \qquad (1)$$

The most active recognition unit from F2 passes a one back to the comparison layer F1. Since the recognition layer is now active, G_1 is inhibits, we set its output to zero for the first digit. In accordance to the rule 2/3, from three different sources two are activated in order to generate an excitatory output. The only comparison units that will fire are those that receive simultaneous ones from the input vector and the recognition layer. The units not receiving a top-down signal from F2 is set inactive even if they receive input from below. This is summarized in Eq. 2.

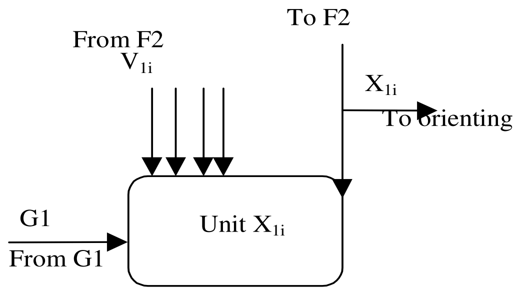

Fig. 3: Processing element X_{1i} in layer F1

$$X_{1i} = \begin{cases} 1 & Ii \cap V_{1i} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

If there is a good match between the top-down template and the input vector, the system becomes stables and the learning start to occur. If there is a mismatch between the input vector and the activity coming from the recognition layer, this indicates that the pattern being returned is not the one desired and the recognition layer should be inhibited.

Search: The reset layer in the orienting subsystem measures the similarity between the input vector and the recognition layer input pattern. If a mismatch between them is found the layer inhibits F2 layer activity. The orienting system compares the input vector to F1 layer output and causes a reset signal if their degree of similarity is less than the vigilance level, where the vigilance parameter set as

$$0<\rho<1 \tag{3}$$

The input pattern mismatch occurs if the following equation is true.

$$\rho \le \frac{|X_1|}{|I|} \tag{4}$$

If the two patterns differ by more than the vigilance parameter, a reset signal is sent to disable the firing unit in the recognition layer F2. The effect of the reset is to force the output of the recognition layer back to zero disabling it for the duration of the current classification in order to search for a better match. The parameter ρ determines how large a mismatch is tolerated or how fine the categories will be before an uncommitted node is chosen. Higher vigilance parameter is used to make the system to search for new categories in response to small difference between I and X_2 learning to classify input patterns into a large number of finer categories.

Learning: The previous three stages take place quickly relative to the time constants of the learning equations of the LTM traces between F1 and F2. The learning process occurs only when the STM reset and search process end all STM parameters on F1 and F2 are stable. The following equation has been used so that the LTM traces from F1 and F2.

$$\tau_1 \frac{dW_{12ij}}{dt} = \begin{cases} (1-W_{12ij})L - W_{12ij}(|X_1|-1) \text{ if } V_{1i} \\ \qquad \text{and } V_{1j} \text{ are active} \\ -|X_1|W_{12ij} \text{ if only } V_{1j} \text{ is active} \\ 0 \qquad \text{if only } V_{1j} \text{ is inactive} \end{cases} \tag{5}$$

where τ_1 is the time constant, X_1 is the processing element in layer F1 as shown in Fig. 3 and L is a parameter with a value greater than one, because the time constant τ_1 is sufficiently larger than the STM activation and small than the input parameter presentation.

The processing element X_{1j} in layer F2 receives input from pattern I_i gain control signal G_2 and V_{2j} equivalent to the outputX_{2i} from F2 multiplied by interconnection weight. The local activity serving also as unit output is X_{1i}. Eq. 5 is the slow learning equation that has been converged in the fast learning equation below.

$$W_{12ij} = \begin{cases} \dfrac{L}{(L-1)+|X_1|} & \text{if } V_{1i} \text{ and } V_{1j} \text{ are active} \\ & \text{if only } V_{1j} \text{ is active} \\ \text{no change} & \text{if only } V_{1j} \text{ is inactive} \end{cases} \tag{6}$$

The initial values for W_{12ji} was randomly chosen while satisfying the following inequality

$$0 < w_{12ij} < \frac{L}{(L-1)+|M|} \tag{7}$$

where M is the input dimension equal to the number of nodes in F1. The LTM traces follows the equation

$$\tau_2 \frac{dW_{12ij}}{dt} = X_{2j}(-W_{12ij} + X_{1i}) \tag{8}$$

where τ_2 is the time constant and the equation is defined to converge during a presentation of input pattern. Thus the fast learning equation of ART1 for W_{12ji} is

$$W_{12ij} = \begin{cases} 1 \text{ if } V_{1i} \text{ and } V_{1j} \text{ are active} \\ 0 \text{ if only } V_{1j} \text{ is inactive} \end{cases} \quad (9)$$

The initial value of the fast learning equation W_{21ij} is chosen to satisfy the inequality.

$$1 \geq W_{12ij} \geq C \quad (10)$$

where C is decided by the slow learning equation (5) para-meter and W_{21ij} is set to 1.

Recognition: In the recognition or bottom-up activation, no input vector is applied in order to disable all recognition in F2 and making the two control gains, G_1 and G_2, equal to zero. This causes all F2 elements to be set to zero, giving them an equal chance to win the subsequent recognition competition.

$$G_1 = \begin{cases} 1 \text{ if } I \neq 0 \text{ and } X_2 = 0 \\ 0 \text{ otherwise} \end{cases} \quad (11)$$

when an input vector is applied one or more of its components must be set to one thereby making both G_1 and G_2 equal to one. Thus, the control gain G_1 depends on both the input vector I and the output X_2 from F2. In other words, if there is an input vector I and F2 that is not actively producing output, then $G_1 = 1$. Any other combination of activity on I and F2 would inhibit the gain control from exciting units on F1. On the other hand, the output G_2 of the gain control module depends only on the input vector I.

$$G_2 = \begin{cases} 1 \text{ if } I \neq 0 \\ 0 \text{ otherwise} \end{cases} \quad (12)$$

If there exists an input vector then $G_2 = 1$ and recognition in F2 is allowed. Each node in F1 receiving a nonzero input value generates an STM pattern activity greater than zero and the nodes output is an exact duplicate of input vector. Since both X_{1i} and I_i are binary, their values would be either 1 or 0, as follows:

$$X_1 = I \text{ if } G_1 = 1 \quad (13)$$

Each node in F1 whose activity is beyond the threshold sends excitatory outputs to the F2 nodes. The F1 output pattern X_1 is multiplied by the LTM traces W_{12} connecting from F1 to F2. Each node in F2 sums up all its LTM gated signals.

$$V_{2j} = \sum_i X_{1i} W_{12ji} \quad (14)$$

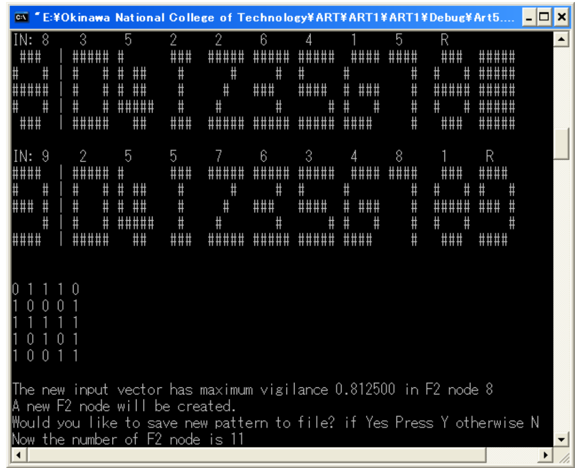

Fig. 4: Simulation output of the ART1

These connections represent the input pattern classification categories, where each weight stores one category. The output X_{2j} is defined so that the element that receives the largest input should be clearly enhanced.

$$X_{2j} = \begin{cases} 1 \text{ if } G_2 = 1 \cap V_{2j} = \max \\ 0 \text{ otherwise} \end{cases} \quad (15)$$

The F2 unit receiving the largest F1 output is the one that best matches the input vector category, thus winning the competition. The F2 winner node fires, having its value set to one, inhibiting all other nodes in the layer resulting in all other nodes being set to zero. After the other nodes are set to zero a new F2 node for a new pattern is created.

Simulation results: The simulation result is shown in Fig. 4 and due to the length of the vertical classification of 0-9digits only the last two[8,9] are shown. As indicated in the introduction, execution time was a major concern in this approach during the simulation test, because we want the ART1 to make a fast learning of the horizontal and vertical classification of 0-9 digits within a short period of time. So the first and fourth layer selectivity values were restricted by two orders of magnitude. Shorten the execution time, restriction of first and fourth layer selectivity values allowed also the classification performance of the network to be readily visualized as a performance surface. The parameter values of the network used in the test are:

- The number of variables in each pattern consists of 5×5 matrix Fig. 4

- The number of nodes in F2 layer is set to 12
- Vigilance value used in the ART1 module is 0.89
- The maximum number of the label used for creating a new F2 node in order to save a new input pattern in it is 999
- The number of patterns stored is 10 and the limiting number of ART1 output nodes is 20
- The learning rate is set to 0.8

In the back-propagation test, a standard back-propagation network is used with no bias units. During the training phase, ART1 is trained on the training data with each module receiving the training data in a different, random and order respectfully. Next we trained again the ART1 to follow the three phases below during the output of the learning result.

- Give first the information on the stored number of input pattern
- Give the number of the variable in each pattern, the vigilance value used and then state the leaning situation in progress. In Fig. 4 we can see that the ART1 has moved the digit 4 and inserted it between 0 and 1 as it has been trained to do in the horizontal classification, while in the vertical classification no digits perturbation has been made. This is to know how far the adaptive resonance theory can be manipulated. R denotes the resonance, IN(example IN: 8) denotes the input and 2,5,5,7,6,3,4,8,1 are the arbitrary classification number used by ART1 as a code of each digit being recognized
- Finally to make sure that no error has been occurred during the learning process, the ART1 should reproduced the number of variable in each pattern (5×5 matrix) for verification, following by the maximum vigilance value of the new input vector F2 node 8, then create a new F2 node Fig. 4. At this stage the user has two choices: Save the new pattern to file or not. If the yes option is chosen then ART1 gives the number of a new node F2. In this test the new node F2 is 11

CONCLUSION

The ART1 based horizontal and vertical classification of 0-9 digits recognition has been modeled using C++ tools. The novelty of this study can be described by the fact that, besides horizontal and vertical recognition, ART1 can move any digit and insert it at any places by assigning to each digit one recognition code during the learning process before output the result within the allocated time. During the matching, F1 node in ART1can remains active only if it receives significant inputs both bottom-up and top-down. In the search procedure, recognition neurons are disabled one by one by the reset function until the vigilance parameter is satisfied by a recognition match.

REFERENCES

1. Grossberg, S., 1976. Adaptive Pattern Classification and Universal Recoding I, I: Feedback Expectation Olfaction and Illusions, BioiCybern, 23: 187.
2. Carpenter, G.A. and S. Grossberg, 1985. Category L earning and Adaptive Pattern Recognition: a neural network model, in proceedings of third army conference on applied mathematics and computing, ARO Report 86-1, 37-56.
3. Carpenter, G.A. and S. Grossberg, 1987. A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine, Computer Vision Graphics Image Processing, pp: 37-54.
4. Grossberg, S., 1987. Competitive learning: From interactive activation to adaptive resonance Cognitive Science (Publication)
5. Carpenter, G.A. and S. Grossberg, 2003. Adaptive Resonance Theory, In M.A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, Second Edn. Cambridge, MA: MIT Press, pp: 87-90.
6. Hussain, T.S. and R.A. Browse, ARTSTAR: A supervised adaptive resonance classifier 10th canadian conference on artificial intelligence (May 16-20, Banff, AB), pp: 121-130.
7. George, M., 2007. Artificial intelligence based three-phase unified power quality conditioner, J. Comput. Sci., 3 (7): 465-477.