Properties of Learning in ART1

Michael Georgiopoulos
Department of Electrical Engineering,
University of Central Florida, Orlando, FL 32816 USA
Gregory L. Heileman
Department of Electrical and Computer Engineering,
University of New Mexico, Albuquerque, NM 87131 USA
Juxin Huang
Department of Electrical Engineering
University of Central Florida, Orlando, FL 32816 USA

Abstract

We consider the ART1 neural network architecture. Useful properties of ART1, associated with the learning of an arbitrary list of binary input patterns, are examined. These properties reveal some of the "good" characteristics of the ART1 neural network architecture when it is used as a tool for the learning of recognition categories. In particular, it is found that if ART1 is repeatedly presented with an arbitrary list of binary input patterns, learning self-stabilizes in at most m list presentations, where m corresponds to the number of distinct size patterns in the input list.

1 Introduction

A neural network architecture for the learning of recognition categories was derived and analyzed by Carpenter and Grossberg in [1]. This architecture was termed ART1 in reference to the *adaptive resonance theory* introduced by Grossberg [2].

It was shown in [1] that ART1 self-organizes and self-stabilizes its recognition codes in response to arbitrary orderings of arbitrarily many and arbitrarily complex binary input patterns. In this article, the timing aspect of this self-organization and self-stabilization process is examined for the fast learning case. Actually, an upper bound is derived on the number of list presentations required by ART1 to self stabilize the recognition codes of an input list of arbitrary binary input patterns that is repeatedly presented to ART1. The upper bound depends only on the number of distinct size binary patterns in the input list. (The size of a binary input pattern is equal to the number of its components that are one.) In particular, if ART1 is repeatedly presented with an input list of m distinct size patterns it will self-stabilize the recognition codes of the input list in at most m list presentations. Furthermore, other useful properties of learning in ART1, induced by its internal dynamics, are described. These properties reveal some of the characteristics of ART1 as a self-organizing neural network architecture for the learning of recognition categories.

2 Model–Preliminaries

A complete description of ART1 and the theorems that give insight into its operation are provided in [1]. The heart of ART1 consists of two interconnected layers of neurons, called the F_1 and F_2 layers. Every node in the F_1 layer is connected via bottom-up traces to all the nodes in the F_2 layer. Input patterns are presented at the F_1 layer. Every node in the F_2 layer is connected with all the nodes in the F_1 layer via top-down traces. The results of the paper are valid under the following assumptions:

A1: All hypotheses of section 18 in [1] hold (one of the hypotheses is that fast learning occurs).

A2: $L - 1 \leq |I|^{-1}$.

A3: $1 \leq |I| \leq M - 1$.

CH 3065-0/91/0000-2671 $1.00 © IEEE

A4: F_2 has at least N nodes.

where $|I|$ is the size of an arbitrary pattern I in the input list, M is the number of nodes in the F_1 layer, N is the number of patterns in the input list and L is a parameter associated with the adaptation of bottom-up and top-down traces in the ART1 neural network architecture.

The top-down traces that emanate from a node in the F_2 layer are called templates. In this paper, only the fast learning case is considered; in that case we let the bottom-up traces and top-down traces reach their limiting values. Initial values for the top-down traces can be taken, without loss of generality, equal to one. To prove our results, the templates of ART1 need to be considered either prior to a pattern's presentation or after a template has fast learned a pattern. Hence, for the purposes of the results discussed in this paper, the ART1 templates can always be thought of as binary vectors.

Consider a pattern I in the list and a template V corresponding to an F_2 layer node. There is a one-to-one correspondence between the components of the binary vectors I and V. A component of I corresponds to a component of V if both of them activate the same F_1 layer node. We define, as in [1], three types of learned templates with respect to an input pattern I: *subset* templates, *superset* templates and *mixed* templates. The components of a subset template V satisfy $V \subseteq I$. They are one only at a subset of the corresponding I components. The components of a superset template V satisfy $V \supset I$. They are one at all the corresponding components of I that are one, as well as at some components of I that are zero. The components of a mixed template V are one at some, but not all of the corresponding I components, as well as at some of the components of I that are zero. In this case, the set of the V components that are one is neither a subset nor a superset of the set of the I components that are one. Sometimes it is convenient to refer to a pattern I as being a subset, superset or mixed pattern with respect to a template V if $I \subseteq V$, $I \supset V$ or V is a mixed template with respect to I. We say that a learned template in the ART1 network is *stable* it it can not be modified by the presentation of any pattern in the input list. Besides the learned templates described above, we also define a template V to be an *uncommitted* template if it corresponds to a node that has not learned any pattern yet. We assume that the components of an uncommitted template consist of all ones.

Since an input pattern I is a binary vector and a template V can be thought of as a binary vector, we define by $|I|$ and $|V|$ the *size* of the binary vectors I and V, respectively. Furthermore, if I is a pattern of the input list and V is a template of ART1, we define $I \cap V$ as the binary vector with ones only at components where both the I and V components are one, and zeros at all other components. Let us now assume that an input pattern I is presented at the F_1 layer. The activity at the F_1 layer changes from 0 to I. Let us also also assume that a node in the F_2 layer with template V is searched at some point during I's presentation. The activity at F_1 changes from I to $I \cap V$. If $|I \cap V| \cdot |I|^{-1} \geq \rho$ we say that the template V *codes* pattern I. If instead $|I \cap V| \cdot |I|^{-1} < \rho$, then the node with template V is *reset* and another node in the F_2 layer is searched. The parameter ρ, called vigilance, determines whether the top-down template is a good match of the input pattern I. We also say, as in [1], that a pattern I has *direct access* to template V if presentation of I leads at once to activation of the F_2 layer node with corresponding template V, and this template codes I on that trial.

Carpenter and Grossberg made in [1] the following conjecture: If the F_2 layer has at least N nodes and if the hypotheses of Section 18 in [1] are valid then, each member of a list of N binary input patterns that is cyclically presented to ART1 will have direct access to an F_2 layer node after at most N list presentations. In this paper, under assumptions A.1-A.4 we prove a much stronger result. The result states that the number of list presentations required by ART1 to learn an arbitrary list of binary input patterns that is repeatedly presented to ART1 is upper bounded by the number of distinct size patterns in the input list. Considering that N is an integer between 2 and 2^M-2 (patterns of size 0 and M are excluded) one can see that in most cases of interest the number of distinct size patterns is much smaller than N. It should be emphasized that our result is stronger than Carpenter and Grossberg's conjecture for ART1 networks that satisfy Assumption A.2. In the case where A.2 is not valid, the aforementioned result does not hold. Consequently, Carpenter and Grossberg's conjecture for ART1 networks that do not satisfy Assumption A.2 is still an open problem.

3 Results

In the following discussion two Lemmas are presented that are instrumental for the proof of our results. Lemma 1 is valid under assumptions A.1-A.3.

Lemma 1 *Suppose that I is an arbitrary pattern from the input list. Learned subset templates with respect to I are searched first in order of decreasing size (i.e., the closest learned subset template to I is searched first, and if it is reset, the next closest subset template to I is searched and so on). If all learned subset templates are reset then, superset and mixed learned templates, as well as uncommitted templates are searched, not necessarily in that order.*

Lemma 1 is a shortened restatement of Theorem 7 in [1] and its proof can be found there. It is obvious by the description of the reset mechanism in Section 2, that if a template V_1 is searched first and reset (i.e., $|I \cap V_1| \cdot |I|^{-1} < \rho$), then any other template V_2 that is searched later will be reset if $|I \cap V_2| \cdot |I|^{-1} \leq |I \cap V_1| \cdot |I|^{-1}$. Lemma 2 is an immediate consequence of Lemma 1 and the above discussion. Lemma 2 is valid under assumptions A.1-A.3.

Lemma 2 *Suppose that I is an arbitrary pattern from the input list, V_1 is a learned subset template (with respect to I) and V_2 is an arbitrary mixed learned template (with respect to I), prior to I's presentation. Then, if V_1 is reset and V_2 is searched, V_2 will also be reset if*

$$\frac{|I \cap V_2|}{|I|} \leq \frac{|I \cap V_1|}{|I|}.$$

Our results are presented in the form of properties (P1–P5) of the ART1 neural network. P1,P2 and P3 address the issue of whether learning in ART1 utilizes network resources wisely. P4 and P5 are the most important properties, because they address the timing aspect of the self-stabilization process in ART1. Properties P1,P2,P3 are valid under assumptions A.1-A.3, while properties P4,P5 are valid under assumptions A.1-A.4.

P1: *In ART1, learned templates are distinct.*

P2: *In ART1, the number of learned templates is smaller than or equal to the number of patterns in the input list.*

P3: *If ART1 is repeatedly presented with an arbitrary list of binary input patterns then, after learning has stabilized, there may exist learned templates which are not directly accessed by any pattern in the input list.*

P4: *Consider an arbitrary list of N binary input patterns that is repeatedly presented to ART1. Then, in list presentations $\geq x$, where $2 \leq x \leq M$:*

- **a)** *A pattern I of size $\geq x$ cannot be coded by a mixed template V, such that $|I \cap V| \leq x - 1$.*
- **b)** *A pattern I of size $\leq x - 1$ will have direct access to a stable template that has been created in list presentations $\leq x - 1$.*

P5: *Consider an arbitrary list of N binary input patterns that is repeatedly presented to ART1, such that k_i $(1 \leq i \leq m)$ of them are of size l_i, where $l_i < l_j$ for $i < j$ $(1 \leq i, j \leq m)$ and $\sum_{i=1}^{m} k_i = N$. Let us denote by S_i $(1 \leq i \leq m)$ the set of input patterns of size l_i. Then, in list presentation $\geq x$, where $2 \leq x \leq m + 1$*

Pattern	First List		
	$V(1)$	$V(2)$	$V(3)$
$I_1 =$ ○○○●●●	○○○●●●		
$I_2 =$ ○○●●○○	○○○●○○		
$I_3 =$ ●●●●○○	○○○●○○	●●●●○○	
$I_4 =$ ○○●●●●	○○○●○○	○○●●○○	
$I_5 =$ ○○○○●●	○○○●○○	○○●●○○	○○○○●●

Figure 1: Template formation in ART1 when the five patterns $I_1 = 000111$, $I_2 = 001100$, $I_3 = 111100$, $I_4 = 001111$, and $I_5 = 000011$ are presented in the order $I_1 I_2 I_3 I_4 I_5$. The vigilance parameter ρ was chosen in the interval $(\frac{1}{4}, \frac{1}{2}]$, and L was chosen according to Assumption A.2.

a) *Every pattern I of the set S_{x-1} has access to a subset learned template V that can code I; V is created in list presentations $\leq x - 1$.*

b) *The presentation of a pattern I from the set S_{x-1} can neither create new templates nor modify already existing learned templates.*

Properties P1,P2 and P5 are proven in [4]. Property P4 is proven in [3]. Let us now try to demonstrate property P3 by presenting two examples. In the example shown in Figure 1, five patterns $I_1 = 000111$, $I_2 = 001100$, $I_3 = 111100$, $I_4 = 001111$, and $I_5 = 000011$ are presented at the F_1 layer of ART1. The order of pattern presentation is kept fixed from list presentation to list presentation. In this figure, the templates $V(1)$, $V(2)$, and $V(3)$, emanating from nodes 1, 2, and 3 in the F_2 layer are also shown. The input patterns and the templates are represented as sequences of open and filled circles—where an open circle stands for the value of zero and a filled circle stands for the value of one. The first row of Figure 1 shows the first pattern presented (i.e., I_1) in the first list presentation and the templates formed after its presentation. The second row of Figure 1 shows the second pattern presented (i.e., I_2) in the first list presentation and the templates formed after its presentation, and so on. Prior to any pattern presentation, the templates $V(1)$, $V(2)$, and $V(3)$ are vectors whose elements consist entirely of ones. In Figure 1, the templates $V(1)$,$V(2)$, and $V(3)$ are shown only if they differ from the vector containing all ones. The vigilance parameter ρ is chosen to be a number in the interval $(\frac{1}{4}, \frac{1}{2}]$, and the parameter L is chosen according to Assumption A.2. After the first list presentation, the templates $V(1) = 000100 = I_1 \cap I_2$, $V(2) = 001100 = I_3 \cap I_4$ and $V(3) = 000011 = I_5$ have been created. Note that the initial values of the bottom-up traces of ART1 have to satisfy certain constraints for this template formation to occur. After the first list presentation learning has stabilized. That is, no new templates are created and already existing learned templates are not modified. In list presentations ≥ 2, patterns I_1 and I_5 are coded by template $V(3)$, patterns I_2 and I_3, are coded by template $V(2)$, and pattern I_4 is coded by either $V(2)$ or $V(3)$ (depending upon the strength of the bottom-up traces converging to nodes 2 and 3 with templates $V(2)$ and $V(3)$, respectively).

The next example, shown in Figure 2, is the same as the one shown in Figure 1, except that the order of pattern presentation is now $I_1 I_3 I_2 I_4 I_5$. The order of this pattern presentation is also kept fixed from list presentation to list presentation. The initial values for the bottom-up traces, top-down traces, the vigilance parameter ρ, and the parameter L are chosen exactly the same as in the previous example. In Figure 2 we see that after the first list presentation, the templates $V(1) = 000011 = I_1 \cap I_5$ and $V(2) = 001100 = I_3 \cap I_2$ have been created. After the first list presentation, learning has also stabilized. That is, no new templates are created and already existing learned templates are not modified. In list presentations ≥ 2, patterns I_1 and I_5 are coded by template $V(1)$, patterns I_2 and I_3 are coded by template $V(2)$, and pattern I_4 is coded by either $V(1)$ or $V(2)$ (depending upon the strength of the bottom-up traces converging to nodes 1 and 2 with templates $V(1)$ and $V(2)$, respectively).

In the previous two examples, ART1 was presented with an identical list of binary input patterns. The only difference being the order of pattern presentation within the list. In Figure 1 pattern I_2

Pattern	First List	
	$V(1)$	$V(2)$
$I_1 =$ ○○○●●●	○○○●●●	
$I_3 =$ ●●●●○○	○○○●●●	●●●●○○
$I_2 =$ ○○●●○○	○○○●●●	○○●●○○
$I_4 =$ ○○●●●●	○○○●●●	○○●●○○
$I_5 =$ ○○○○●●	○○○○●●	○○●●○○

Figure 2: Template formation in ART1 when the five patterns $I_1 = 000111$, $I_2 = 001100$, $I_3 = 111100$, $I_4 = 001111$, and $I_5 = 000011$ are presented in the order $I_1 I_3 I_2 I_4 I_5$. The vigilance parameter ρ was chosen in the interval $(\frac{1}{4}, \frac{1}{2}]$, and L was chosen according to Assumption A.2.

is presented second and pattern I_3 is presented third, while in Figure 2 the order of presentation for patterns I_2 and I_3 is reversed; the other input patterns are presented with the same order in the list for both examples. We observe from these examples that even a single change in the order of pattern presentation within a list can lead to the formation of different templates. Carpenter and Grossberg have also demonstrated this characteristic of ART1 in the computer simulation results presented in [1]. A more interesting observation can be derived by examining the templates formed in the example shown in Figure 1. In this example, after learning has stabilized, we observe that the learned template $V(1)$ is not directly accessed by any pattern in the input list. This behavior is not exhibited in the example shown in Figure 2. The examples in Figures 1 and 2 verify the validity of property P3.

It is worth noting that Carpenter and Grossberg have been very careful about the issue of direct access. In [1] they prove Theorem 6 which states that after learning has stabilized in ART1 networks satisfying Assumption A.1, each input pattern either has direct access to a learned template in the F_2 layer, or the input pattern cannot be coded by any learned template in the F_2 layer. They do not state that every learned template in the F_2 layer, after learning stabilizes, will be directly accessed by at least one pattern in the input list. Property P3 demonstrates that this is not necessarily true.

4 Remarks–Conclusions

In this paper, properties of learning in the ART1 neural network architecture have been presented, under Assumptions A.1-A.4. Their proof can be found in [3] and [4].

Property P4 states that if an ART1 network is repeatedly presented with an arbitrary list of binary input patterns it self-stabilizes the recognition codes (templates) of size-l ($1 \le l \le M-1$) patterns in at most l list presentations. Property P5 states that if ART1 is repeatedly presented with an arbitrary list of binary input patterns learning stabilizes in at most m list presentations, where m is the number of distinct size patterns in the input list. Since our modeling assumptions exclude patterns of size 0 or M, m ranges from 1 to $M-1$. In most practical applications it might be difficult to examine the patterns in the input list in order to determine the number of distinct size patterns. In most cases though, we know the number N of patterns in the input list, as well as the number of nodes M in the F_1 layer of ART1. In these cases we can say that learning in ART1 stabilizes in at most $\min(N, M-1)$ list presentations. Another interesting observation, originating from part b of property P5 is that the template formation and learning in ART1 will not be affected if patterns in the set S_x ($x \ge 1$) are presented only in the first x list presentations.

Properties P1 and P2 are also of particular interest because they show that learning in ART1 does not waste network resources by creating identical templates or by creating more templates than the number of patterns that we try to categorize. Nevertheless, property P3 indicates that there might be a possibility, after learning has stabilized, that some templates are not directly accessed by any pattern in the input list.

It is worth observing that properties P1,P2,P4,P5 are valid independent of the order in which the input patterns are presented within the list. In addition, the ordering of the patterns within the list can change from one list presentation to the next without affecting the validity of these properties.

Acknowledgements

The authors would like to thank Professor Gail Carpenter for her thoughtful suggestions regarding the contents of the papers in references [3] and [4]. This research was supported by a grant from the Florida High Technology and Industry Council with matching support from Martin Marietta Electronic Systems Division.

References

[1] G. A. Carpenter and S. Grossberg, "A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine", *Computer Vision, Graphics, and Image Processing*, **37**, 54–115, 1987.

[2] S. Grossberg, "Adaptive Pattern Recognition and Universal Recoding, II: Feedback, Expectation, Olfaction, and Illusions", *Biological Cybernetics*, **23**, 187–202, 1976.

[3] M. Georgiopoulos, G. L. Heileman, and J. Huang, "Convergence Properties of Learning in ART1," *Neural Computation*, Vol.2, No.4, Winter 1990, pp. 502-509.

[4] M. Georgiopoulos, G. L. Heileman, and J. Huang, "Properties of Learning Related to Pattern Diversity in ART1," *Neural Networks*, Vol.4, No. 6, November 1991; to appear.