ADAPTIVE CLUSTERING BY ART2 NEURAL NETWORKS

Anatoli Nachev
Management Information Systems, Dept. Accountancy &Finance, National University of Ireland, Galway, Ireland
E-mail: Anatoli.Nachev@nuigalway.ie

Ivan Ganchev
Dept. of Electronic & Computer Engineering, University of Limerick, Limerick, Ireland
E-mail: Ivan.Ganchev@ul.ie

ABSTRACT

This paper proposes a general learning mechanism for ART2 neural networks that relaxes the specification within the classic ART2 model, which restricts learning only to the active node. Thus the learning allows slow forgetting by exponential decay of all long-term memory traces. This approach changes the ATR2 learning rules, but does not use additional node features, or supervisory subsystems. It preserves the basic ART2 architecture and functioning, whereby makes the implementation straightforward. The proposed general learning mechanism releases redundant committed nodes for further learning, helps to prevent the system from blocking, and enhances a variety of network features. It may be used for some classes of applications that require clustering in a very large input space, or rapidly changing environmental conditions.

1 INTRODUCTION

Neural modelling, based on the adaptive resonance theory (ART), clarifies how sensory and cognitive processes solve a key problem, called *stability – plasticity dilemma* [1], whereby the brain can rapidly learn about the world throughout life without catastrophically forgetting our previous experiences. In other words, we remain plastic and open to new experiences without risking the stability of previously learned memories. This type of fast stable learning enables us to dealing in some degree with changing environmental conditions. Old knowledge representations can be refined by changing contingencies, and new ones built up, without destroying the old ones due to catastrophic forgetting. On the other hand, catastrophic forgetting is a good property for spatial and motor learning [1]. We have no need to remember all the spatial and motor representations that we used when we were children. In fact, the parameters that controlled our small childhood limbs in space would cause major problems if they continued to control our larger and stronger adult limbs.

The forgetting phenomenon has been modelled by many modifications of artificial neural network. A method of training and zeroing of fuzzy neural networks (FNN) [6] use a technique, which zeroes weak connections before training the network with new data. Another technique uses decay of connection weights during training of FNN, followed by pruning (cutting off) weak connections [7]. A modification of the self-organising feature maps of Kohonen erases redundant information by dropping out nodes [5], [10]. Most of the models of forgetting, however, incorporate additional node features or additional supervision of the learning process.

The classic ART2 model of neural networks solves in a satisfactory manner the problem stability – plasticity, however it demonstrates some drawbacks when used in some classes of applications. For example, in case of rapidly changing environmental conditions, when the network cannot classify some of the new incoming input patterns into existing categories, it creates new ones. In this case the network commits uncommitted nodes instead of reusing those, which contain old and not actual data.

The general learning of ART2 neural networks, discussed in this paper, models an additional mechanism of forgetting, designed to reuse those committed nodes, which contain useless old data, thus making space for new one. The proposed approach does not use additional node features, or supervisory subsystems. It changes the learning rules, but preserves the basic ART2 architecture and functioning, whereby makes the implementation straightforward.

2 GENERAL ART2 LEARNING

A typical ART2 architecture consists of several components [2], illustrated in Figure 1. The input representation field F1 is composed by three subfields of nodes, which act as a short-term memory (STM) of the system. The category representation field F2 contains top-down and bottom-up long-term memory (LTM) traces that store learned categories. The orienting subsystem compares a presented input pattern with a category.

2.1 Classic ART2 Functioning

An incoming input pattern first goes through transformations in the field F1, and then activates a competition between all nodes in the field F2, each of which represents a stored category. The goal of this competition is to find a node, which matches with the presented input pattern better than all other competitors. The node, which wins the competition, becomes active. It provides the orienting subsystem with a category pattern in order to be matched with the presented input pattern. The orienting subsystem estimates the similarity between them calculating the length of the vector *r*. In case it exceeds a threshold (network parameter of vigilance), the system gets in resonance state. Then it starts learning of the active node, which adjusts its category pattern to be closer to the input pattern. If the matching in the orienting subsystem fails, a reset signal suppresses the active node and a new competition without all suppressed nodes takes place. Repeating this process, the neural network either finds a category that matches well with the input pattern, or an uncommitted node learns the input pattern creating a new category.

In both cases the learning involves only one node – either the active one or an uncommitted one. The ART2 model does not involve passive nodes in learning in order to preserve the rest of learned categories from catastrophic forgetting.

Differential equations (1) and (2) represent the learning rules of the classic ART2 model [2]. The first one describes how the top-down LTM traces change, whereas the second one describes the change of bottom-up LTM traces.

Top-Down ($F2 \rightarrow F1$):

$$\frac{d\,z_{ji}}{dt} = g(y_j)(p_i - z_{ji}) \qquad (1)$$

Bottom-Up ($F1 \rightarrow F2$):

$$\frac{dz_{ij}}{dt} = g(y_j)(p_i - z_{ij}) \qquad (2)$$

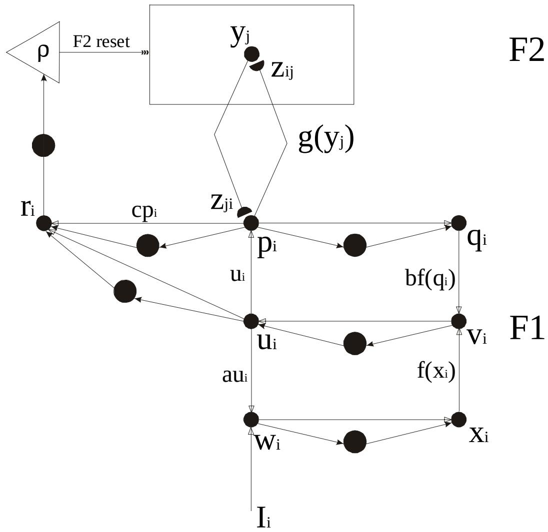

Figure 1: ART2 architecture

2.2 General ART2 Learning Rules

The modification of ART2 model, discussed in this paper, aims to provide an ART2 system with additional flexibility and adaptation to complex or variable input domains, making the network resources reusable. In order to achieve these goals, the modification incorporates a mechanism of *general learning*, which relaxes the specification within the classic ART2 model [2] that restricts learning only to the active node. The general learning mechanism involves both active and passive F2 nodes in learning by replacing the learning rules (1) and (2) of the classic ART2 model with new once, represented by differential equations (3) and (4). Equation (3) represents change of top-down LTM traces, whereas (4) represents change of bottom-up LTM traces .

$$\frac{dz_{ji}}{dt} = g(y_j)[p_i - z_{ji}] - \lambda[d - g(y_j)]z_{ji} \qquad (3)$$

$$\frac{dz_{ij}}{dt} = g(y_j)[p_i - z_{ij}] - \lambda[d - g(y_j)]z_{ij} \qquad (4)$$

The coefficient λ, where $0 \le \lambda < 1$, is a new network parameter which can be used to adjust rate of forgetting. When $\lambda = 0$, (3) and (4) become equivalent to (1) and (2), which corresponds to the classic ART2 learning without forgetting. Obviously the general learning rules can be considered as a generalization of the classic ones.

After equivalent transformations equations (3) and (4) can be represented by (5) and (6) .

$$\frac{dz_{ji}}{dt} = \begin{cases} g(y_j)[p_i - z_{ji}], & if\, g(y_j) = d \\ -\lambda dz_{ji}, & if\, g(y_j) = 0 \end{cases} \tag{5}$$

$$\frac{dz_{ij}}{dt} = \begin{cases} g(y_j)[p_i - z_{ij}], & if\, g(y_j) = d \\ -\lambda dz_{ij}, & if\, g(y_j) = 0 \end{cases} \tag{6}$$

Equations (5) and (6) show that learning occurs even when $g(y_j) = 0$. This means that all F2 nodes, without exception, go through learning every time when the system initiates that process. Learning, however, is different for the active and passive nodes. The active node learns according to the rules, represented by the upper branches of (5) and (6), i.e. applying the classic ART2 model. All passive nodes, however, learn according to the lower branches of (5) and (6), which can be generalised by differential equation (7)

$$\frac{dz}{dt} = -\lambda d.z \tag{7}$$

3 LEARNING VS FORGETTING

Several issues emerge from using the learning rules (5) and (6) in the ART2 architecture.

3.1 Exponential Decay of LTM Traces

We can consider a learning of a F2 node, which takes place in a time period $[t_0, t_1]$. Inequalities $0 < \lambda < 1,\ 0 < d < 1,\ 0 \geq z_{ij},\ 0 \geq z_{ji}$, and equation (7) imply that for this time period inequalities (8) and (9) are valid.

$$\frac{dz_{ij}}{dt} < 0 \tag{8}$$

$$\frac{dz_{ji}}{dt} < 0 \tag{9}$$

Inequalities (8) and (9) show that in this time period all passive nodes decrease their LTM traces.

Differential equation (7) can be solved with initial condition (10).

$$z(t_0) = z^0 \tag{10}$$

The solution can be represented by equation (11).

$$z(t) = z^0 e^{(-\lambda d)(t-t_0)} \qquad t \in [t_0, t_1] \tag{11}$$

Equation (11) implies that for $\lambda \neq 0$ each inactive LTM trace decreases exponentially. The learning that takes place over *k* pattern presentations can be viewed as a series of *k* discrete time slices $[t_0, t_1]$, $[t_2, t_3]$ … $[t_{2k-2}, t_{2k-1}]$. In each time slice the active node learns the input pattern, while the passive nodes learn their own pattern, but weakened. At the end of each time slice an LTM trace has a final value that appears to be an initial one for the next slice. For simplicity, time slices can be considered as a continuous interval $[t_0, t_k]$, ignoring time between learning. The implication is that the LTM traces will reach zero when the length of interval (or number of subintervals) reaches infinity, i.e. $\lim_{t_k \to \infty} z(t) = 0$. Therefore during learning the inactive LTM traces tend to zero, but never reach it.

The weakening of bottom-up LTM traces impacts on the network behaviour. If a presented input pattern is similar or identical to a weakened category, it is likely that it will not win the competition from the first attempt. Such behaviour resembles the process that leads to forgetting in biological neural systems. Likewise, our own experience shows that if a part of our knowledge is not used for a long time, it ‘goes’ deeply in the mind, making its retrieval harder. When we identify a stimulus, an image perhaps, we first access appropriate and prominent, i.e. common or frequently used categories, and only if this first pass fails we start to access less frequently used, weaker, but still formed traces, and finally if this fails resort to learning, i.e. establishing a new category.

3.2 System Stability

Another issue of the general ART2 learning mechanism is that it preserves stored information during weakening. Decreased LTM traces do not change the shape of the stored categories. Let us view top-down and bottom-up category patterns z_i and z_j of a passive node as multidimensional vectors with components

LTM traces. It can be proven that the weakening makes them shorter, but it preserves their direction. Indeed, we can calculate ratio between two arbitrary chosen traces $\hat{z}$ and $\tilde{z}$, that belong to a passive category pattern, during weakening in a time interval $[t_p, t_q]$. Equation (11) implies that for any $t \in [t_p, t_q]$

$$\frac{\hat{z}(t)}{\tilde{z}(t)} = \frac{\hat{z}^0 e^{(-\lambda d)(t-t_p)}}{\tilde{z}^0 e^{(-\lambda d)(t-t_p)}} = \frac{\hat{z}^0}{\tilde{z}^0} = const$$

This equation shows that the ratio is a constant, which does not depend on the time. The implication is that the weakening preserves the vector direction, therefore encoded information.

3.3 Reusability of Resources

The process of LTM weakening facilitates the 'recycling' of committed nodes that contain useless information. The general learning mechanism releases and reuses such nodes.

It can be proven [3], that applying the general learning rules (3) and (4) to a passive F2 node, there is a finite time interval of weakening, after which the category reaches a critical state, which releases the node. Then it can learn another category, i.e. to be reused.

Indeed, the orienting subsystem is assumed to reset F2 whenever an input pattern is active and $\frac{\rho}{\|r\|} > 1$. From other hand length of the vector *r* during the weakening can be represented by (12).

$$\lim_{t, t_k \to \infty} \|r(t)\| = \frac{[(1+c)^2]^{\frac{1}{2}}}{1+(c^2)^{\frac{1}{2}}} = 1 \qquad (12)$$

This equation implies that for a fixed ρ, where $0 < \rho < 1$, it exists a moment $\tilde{t}$, $t_0 \prec \tilde{t}$, such that for any $t' : \tilde{t} \prec t'$

$$\|r(t')\| > \rho \qquad (13)$$

Equation (13) suggests that after a finite time interval the weakened category has the same characteristics as an empty one. It always causes resonance state for any presented input pattern, and then learns it.

The general learning mechanism affects also the uncommitted nodes by continually weakening, since they are always passive. Despite that the uncommitted nodes remain useable, because the only requirement for their values, represented by (14) [2], is valid during the weakening.

$$0 < z_{ij}(0) \le \frac{1}{(1-d)\sqrt{M}} \qquad (14)$$

3.4 Biological Plausibility

Although biological plausibility of the discussed forgetting mechanism is not main objective of its design, the dynamics and properties that it shows seem to be similar to several basic characteristics of phenomenon forgetting in biological neural systems.

There are many different perspectives, physiological and psychological, from which forgetting can be viewed. Ebbinghaus [8] describes dynamics of the forgetting by a curve that depicts number of memorized items as a function of the time. That curve shows a negative acceleration of the forgetting along the time, which can be explained by the fact that most of the forgetting takes place in the first several hours after the learning. Later it slows down significantly. Kintsch [9] suggest that the curve of Ebbinghaus can be approximated well by function (15).

$$Y(t) = ab^{-t} \qquad (15)$$

Here $Y(t)$ is the quantity of memorised information, represented as a function of the time. It depends on two positive constants a and b. After equivalent transformations, equation (15) can be represented by (16) [4].

$$Y(t) = a_0 e^{-a_1 t} \qquad (16)$$

Equation (16) suggests that the forgetting in biological neural systems can be represented by an exponentially decreasing function of the time, as the general ART2 learning suggests.

Other theories of forgetting in the cognitive psychology provide similar conclusions. For example the multi-component trace theory of Bower [10] offers a function of memorizing $r(t)$, represented by (17).

$$r(t) = J + (1-J)a^t \qquad (17)$$

Here J and a are positive constants. After equivalent transformation [4] (17) can be represented by (18),

$$r(t) = J + C_1 e^{-C_2 t} \quad (18)$$

where the parameters J, C_1, and C_2 are also positive constants. Equation (18) suggests again that the forgetting in biological neural systems can be represented by an exponentially decreasing function of the time, as the general ART2 learning suggests.

3.5 System Flexibility and Adaptation

In ART2 neural networks with general learning the forgetting may be conceived of as a corollary of the release of atrophied or unused resources. It can be seen as a sensible strategy directed at the overall management of specific and limited computational resources. Without forgetting, plasticity and adaptation the neural network may become cluttered with relatively unused resources. For some applications, planning what resources a network needs is impossible and limited resources impose the need to release unused resources for further use.

Neural networks are often given some initial, well-defined sample of instances that represent the domain the network is applied to. The network is expected to learn this training set. Also, there are applications, which are not well defined and for which planning of network resources is more difficult or even impossible, because the network needs to learn about an unknown input space. In these circumstances the system's future inputs are unknown. Often such applications use unsupervised self-organizing learning paradigms like those of ART neural networks.

In a situation where the input space is huge the network may not be able to learn all details required for an adequate functioning of the system. For example if at some arbitrary level of vigilance an classic ART2 networks runs out of resources the system will block, and either the granularity of the categories must be increased with an accompanying loss of detail, or more resources must be allocated. This situation may also arise when the input space is not huge, but continually changing. In this situation as learning continues, not only may more category nodes be required, but existing category exemplars may also shift and some may become unused as their input s become assimilated to other similar categories.

Discussed above general ART2 learning is a possible solution. It allows the network to work continuously without blockage, or unlimited expansion of resources since it releases those nodes, which contain inactive information. Such a solution is attractive because if implemented adequately within the dynamics of an ART2 network the forgotten information could be restored. If patterns that were mapped to the forgotten nodes reappear in the environment they can be relearned.

4 BENCHMARKS

The architecture of an ART2 neural network with general learning rules has been implemented by a simulator, which uses the classic ART2 architecture with pre-processing layer. The learning rules were implemented by solving the differential equations (3) and (4) using the fourth order Runge-Kutta method. The simulator was adjusted for stable performance by the network parameters as follows: $\rho = 0.98$, $a = b = 10$, $d = 0.9$, $\theta = 0.17$. Parameter λ $0 \le \lambda < 1$, which adjusts the rate of forgetting, used the following values for different simulations: $\lambda = 0; 0.0018; 0.0020; 0.0022$.

The general ART2 learning mechanism was tested by two groups of simulations using several sets of input patterns. The basic one A consisted of 1500 input patterns, each of which was a N-dimensional vector (N=50) of randomly chosen numbers between 0 and 1.

The first group of simulations compared classic ART2 learning with the general one in case where the input set contains additional patterns, capable to establish new categories of 'useless' knowledge. The simulations used an additional input set B of 500 patterns, which were:

- Deliberately damaged samples of A. The origin of such patterns might be practical failure or redundancy of information caused by bad transmission or errors in formation input patterns.
- Patterns, which were presented relatively infrequently to the network. Again these may cause establishment of infrequently used categories. The origin of such patterns might be a wrong learning or learning unimportant for the practice information.

The simulations comprised 250500 random presentations of patterns from both A and B. Each of the input patterns of B was presented after a series of 500 arbitrary chosen patterns form A. The simulations were carried out four times – once with parameter $\lambda = 0$, which corresponds to the classic ART2 learning, and three times with values $\lambda = 0.0018; 0.0020; 0.0022$, which correspond to the general learning with different rates of slow forgetting. After all simulations the input set A was presented to the network in order to determine way of clustering, number of the learned categories, and number of committed nodes. All simulations showed identical clustering of A into 126 categories, but different number of committed nodes. Table 1 shows the results from the first group of simulations.

Model of ATR2 learning	λ	**Number of presentations**	**Established categories for** A	**Committed nodes**
Classic	0	250500	126	247
General	0.0018	250500	126	163
General	0.0020	250500	126	154
General	0.0022	250500	126	138

Table 1 Number of committed nodes used by classic ART2 learning and general ART2 learning after infrequent presentation of untypical input patterns.

Model of ART2 learning	λ	**Number of presentations**	**Established categories for** D	**Committed nodes**
Classic	0	375750	73	129
General	0.0018	375750	73	121
General	0.0020	375750	73	114
General	0.0022	375750	73	101

Table 2 A comparison of committed nosed where the network were presented with a continually shifting input space.

The second group of simulations used two sets of input patterns C and D, which were derived from A. They divided A into two halves by arbitrary extraction of input patterns form A. Each of the halves contained 750 input patterns. The simulations aimed at observing the two learning mechanisms response to a continually changing input space. The initial input set C was gradually changed to D by replacing patterns in C with those in D. After 500 pattern presentations one pattern from D replaces one pattern from C until the input set C became D. The simulations were carried out four times – once with parameter $\lambda = 0$, which corresponds to the classic ART2 learning, and three times with values $\lambda = 0.0018; 0.0020; 0.0022$, which correspond to the general learning with different rates of slow forgetting. Table 2 shows the results from the second group of simulations.

The simulations classified identically the input space D into 73 categories, but using different number of committed nodes.

5 CONCLUSION

This paper discusses a general learning mechanism that allows forgetting within ART2 neural networks. The proposed approach is straightforward and arguably biological plausible. It preserves the main properties of the ART2 architecture, but enhances a variety of network features, releases redundant resources for further learning, and helps to prevent the system from blocking. This model increases the system's ability to drop out error information obtained during learning, and to adapt to a continually changing or very large input spaces.

REFERENCES

[1] S. Grossberg, "Linking Mind to Brain: The Mathematics of Biological Intelligence," Notices of the American Mathematical Society, vol. 47, pp. 1361-1372, 2000.

[2] G. Carpenter and S. Grossberg, "ART2: Self-Organisation of Stable Category Recognition for Analog Input Patterns," Applied Optics, vol. 26, pp. 4916-4930, 1987.

[3] A. Nachev, N. Griffith, A. Gerov "Forgetting in ART2 Neural Networks," Proceedings of the International Conference on Neural Network and Brain ICNN&B, Beijing, China, June, 1998.

[4] A. Nachev, D. Chelebieva, and A. Gerov, "Self-regulation of LTM Trace Decay in Adaprive ART2 Neural Networks," Proceedings of the XXIX Conference of UMB, Sofia, Bulgaria, April, 2000.

[5] B. Fritzke, "Unsupervised Clustering with Growing Cell Structures," presented at the IEEE Conference IJCNN, Seattle, 1991.
[6] N. Kasabov, "Investigating the Adaptation and Forgetting in Fuzzy Neural Networks Through a method of Training and Zeroing," Proceedings of the International Conference on Neural Networks ICNN'96, 1996.
[7] M. Ishikawa, "Neural Network Approach to Rule Extraction," Proceedings of the Second New Zealand International Conference on Artificial Neural Networks and Expert Systems ANNES, IEEE Computer Society Press, 1995 ,pp.6-9.
[8] H. Ebbinghaus, Memory: A Contribution of Experimental Psychology, New York, Dover, 1964.
[9] W. Kintsch, Learning, Memory, and Conceptual Processes, John Wiley & Sons, New York, 1970, pp. 139-143.
[10] H. Keuchel, E. Putcamer, U. Zimmer, "Learning and Forgetting Surface Classification with Dynamic Neural Networks," Proceedings of ICAAN'93, 1993.