Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 21 (1997), 117–129.

CATEGORY ART: A VARIATION ON ADAPTIVE RESONANCE THEORY NEURAL NETWORKS

David Weenink

Abstract

In this paper we describe Category ART, a variation on the adaptive resonance theory (ART) neural network models. Category ART is a predictive ART architecture because it incorporates an ART module to be able to learn to predict a prescribed category given a prescribed n-dimensional input vector **a**. In contrast to ARTMAP, Category ART contains only one ART module and the map field algorithm has been simplified. The remaining ART module in a Categoty ART can be either Fuzzy ART or ART2-A. Its performance is demonstrated on a benchmark neural network test, the two spiraals problem.

1. Introduction

Category ART is a neural network topology whose dynamics are based on Adaptive Resonance Theory (ART). ART was developed by Grossberg (1976a, 1986) as a theory of human cognitive information processing. It was the result of an attempt to understand how biological systems are capable of retaining plasticity throughout life, without compromising the stability of previously learned patterns. Somehow biologically based learning mechanisms must be able to guard stored memories against transient changes, while retaining plasticity to learn novel events in the environment. This tradeoff between continued learning and buffering of old memories has been called by Grossberg the *stability-plasticity dilemma.* This poses special design problems, since, for example, in (supervised) feedforward networks, which are the most popular neural networks nowadays, new information gradually washes away old information, and therefore, feedforward networks cannot be made stable in a changing environment.

To be able to mimic biological behaviour, the emphasis of ART neural networks lies at *unsupervised learning* and *self-organization.* Unsupervised learning means that the network learns the significant patterns on the basis of the inputs only, there is no feedback. There is no external teacher that instructs the network to which category a certain input belongs. Other types of learning are *reinforcement* learning and *supervised* learning. In reinforcement learning the net receives only limited feedback, like "on this input you performed well" or "on this input you have made an error". In supervised mode a net receives for each input the correct response. Unsupervised learning is the substrate on which the other types of learning are based. Learning in biological systems always starts as unsupervised learning, for the newly born hardly any pre-existing categories exist. A system that can learn in unsupervised mode can always be adjusted to learn in the other modes, like reinforcement mode or supervised mode. However, a system specifically designed to learn in supervised mode can never perform in unsupervised mode. Needless to say that in unsupervised mode we cannot have a separate training and performance phase because this implies the presence of a homunculus that knows when to alter phases. Self-organization means that the system must be able to build stable recognition categories in real-time.

These design constraints have led to a series of real-time ART neural network models for unsupervised category learning and pattern recognition. Model families include ART1, which can stably learn to categorize binary inputs presented in an arbitrary order (Carpenter & Grossberg, 1987b); ART 2, which can stably learn to

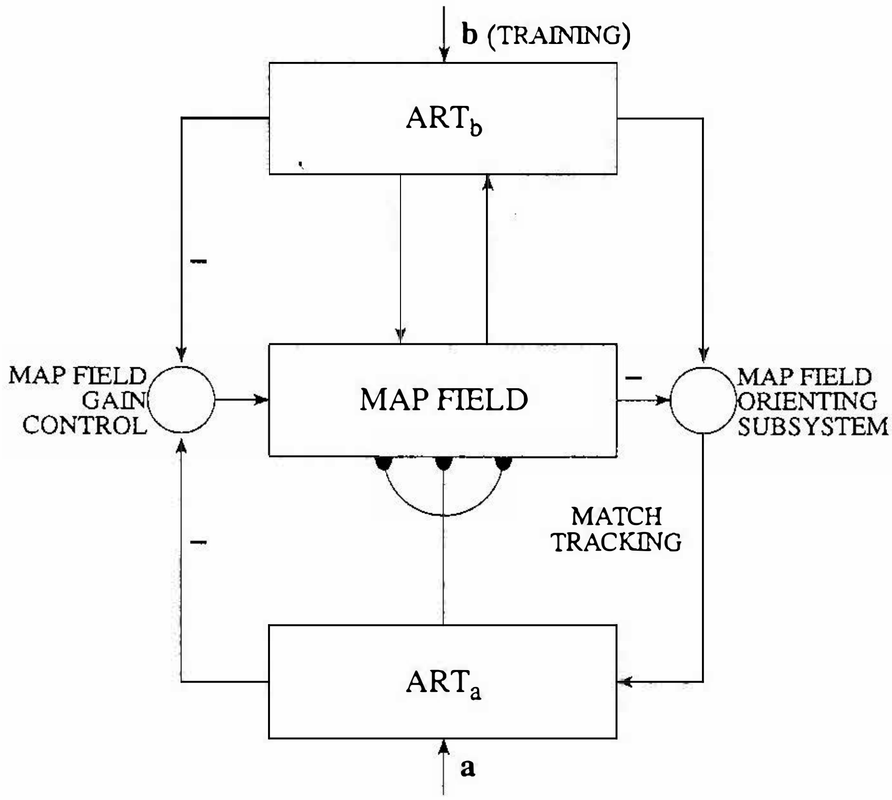

Fig. 1. Block diagram of a supervised ARTMAP system. Two ART modules are linked by an inter-ART module called the map field. The map field forms predictive associations between categories of the ART modules and realizes a match tracking rule. If ART_a and ART_b were disconnected each module would self-organize category groupings for their respective input sets.

categorize either analog or binary data (Carpenter & Grossberg, 1987) and ART 3, which can carry out parallel search of distributed recognition codes in a multilevel network hierarchy (Carpenter & Grossberg, 1990). The Fuzzy ART model (Carpenter et al., 1991) is based on fuzzy logic computations and incorporates the ART1 model since computations from fuzzy set theory reduce to binary computations when the fuzzy variables become binary valued.

Besides the networks described above, which are based on unsupervised learning, *supervised* network architectures like ARTMAP have been developed that incorporate one or more of the unsupervised ART modules given above (Carpenter et al., 1991a). Figure 1 shows a block diagram of such a system. In supervised mode mappings are learned between input vectors **a** and **b**. A familiar example of supervised neural networks are feedforward networks with backpropagation of errors (BP networks, Weenink, 1992). Supervision is, however, their only similarity with ARTMAP networks. ARTMAP networks are *self-stabilizing* while in BP networks new information gradually washes away old information. A consequence of this is that a BP network has separate training and performance phases while ARTMAP systems perform and learn at the same time. Besides, ARTMAP networks are designed to work in *real-time*, while BP networks typically are designed to work off-line, at least during their training phase. Another difference is that while ARTMAP systems can learn both in a fast as well as in a slow match configuration, BP networks can only learn in slow mismatch configuration. This means that an ARTMAP system learns, i.e., adapts its weights, only when the input *matches* an established category, while BP networks learn when the input does *not* match an established category. In BP networks there is always the danger of the system getting trapped in a local minimum while this is impossible for ART systems. However, in systems based on ART modules learning may depend upon the ordering of the input patterns.

Category ART, that we herewith introduce, is a specialized fast *algorithmic* variant of the ARTMAP class of neural network architectures and performs incremental supervised learning of recognition categories in response to input vectors presented in arbitrary order. Under supervised learning conditions, Category ART's internal control machanisms create stable recognition categories by maximizing predictive generalization while minimizing predictive error, just like the ARTMAP architectures do.

Category ART differs from the figure 1 ARTMAP architecture in several ways: there is only one ART module present and the map field has disappeared. Instead a simpler algorithm replaces the dynamics of both components. The dynamics of the network, however, are still based on Adaptive Resonance Theory .

Originally all learning equations in ART systems are written in the language of real-time systems, i.e., differential equations. In our implementation, as in most algorithmic variants discussed above, steady state approximations are used that capture the essence of these dynamic equations. Hence we do not have to use integration methods nor will we use differential equations in the formulation of the dynamics of Category ART.

2. Basic features of ART systems

The basic features of Adaptive Resonance Theory and its relation to perception are layed out in a great number of articles by Grossberg and his associates (see for example Grossberg 1986 for an overview). A block diagram for a typical ART system is displayed in figure 2. The main components are the *attentional* subsystem and the *orienting* subsystem. The attentional subsytem consists, among others, of two fields of neurons, F_1 and F_2, where each field may consist of several layers of neurons. These fields are connected with feedforward and feedback connection weights. The connection weights form the long term memory (LTM) components of the system and multiply the signals along these pathways. The name *short term memory* (STM) will be associated with the pattern of activity that develops on a field as an input pattern is processed. The orienting subsystem is necessary to stabilize the processing of STM and the learning in LTM. As can be seen from the figure, the F_1 field receives input from possibly three sources. These three input sources are the bottom-up input to F_1, the top-down input from F_2 and the gain control signal. To avoid the possibility that mere feedback from F_2 can generate spontaneous activity at level F_1, i.e., to avoid that the system hallucinates, system dynamics are limited in such a way that at least two out of three inputs must be active to generate activity at the F_1 field. This is called the 2/3 rule in ART. The same rule applies to the three possible input sources for the F_2 level.

All ART systems incorporate basic features, notably, pattern matching between bottum-up input and top-down learned prototype vectors. This matching leads either to a resonant state that focusses attention and triggers stable prototype learning or to a

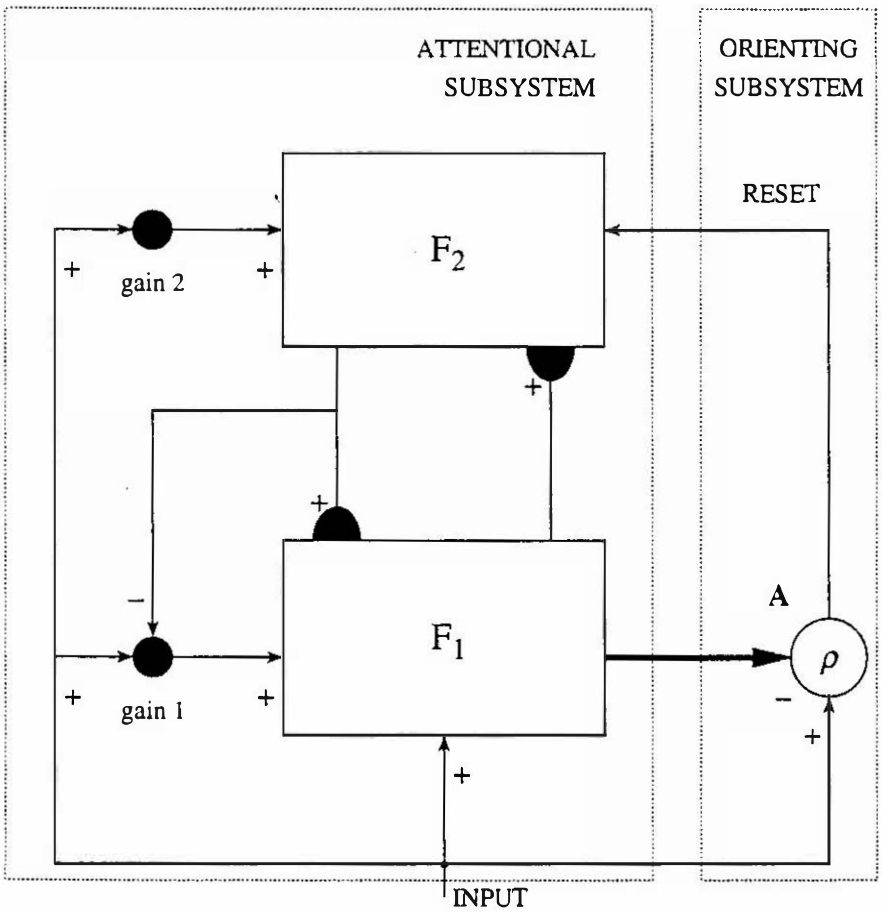

Figure 2. Typical ART neural network block diagram. After preprocessing, the input activity pattern is transformed to the first field F_1. Field F_1 is connected to field F_2 with feedforward and feedback connections which are indicated with black half ellipses. These connections form the long term memory components of this system.

self-regulating parallel memory search. This search ends in either of two ways. First, if an established category is selected, then this prototype may be refined to incorporate new information in the input pattern. In this case when an input matches an established category, we speak of *resonance*. This resonant state persists long enough for learning to occur; hence the term *adaptive resonance theory*. Second, if the search ends by selecting a previously untrained node, then learning of a new category takes place. The criterion of an acceptable match is defined by a dimensionless parameter ρ called *vigilance*. Vigilance weighs how close an input must be to the top-down prototype for resonance to occur. Because the vigilance

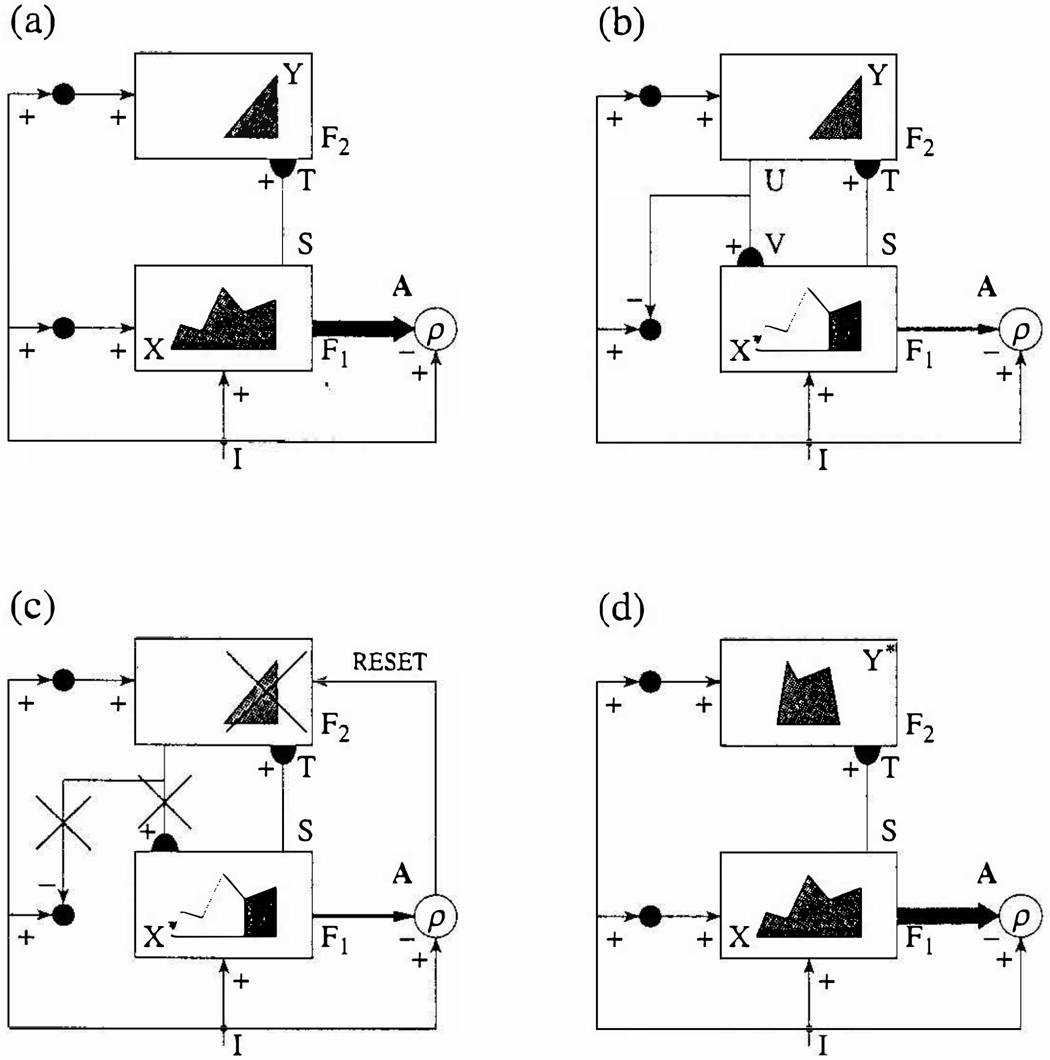

Fig. 3. ART search for an F_2 code: (a) The input pattern I generates, after being properly normalized, the specific STM activity X at F_1 as it nonspecifically activates the orienting subsystem A. Pattern X both inhibits A and generates the output signal pattern S. Signal pattern S is transformed into the input pattern T, which activates the STM pattern Y across F_2; (b) Pattern Y generates the top-down signal pattern U, which is transformed into the prototype pattern V. If V mismatches I at F_1, then a new STM activity pattern X^* is generated at F_1. The reduction in total STM activity which occurs when X is transformed into X^* causes a decrease in the total inhibition from F_1 to A; (c) If the matching criterion ρ fails to be met, A releases a nonspecific arousal wave to F_2, which resets the STM pattern Y at F_2; (d) After Y is inhibited, its top-down prototype signal is eliminated, and X can be reinstated at F_1. Once again X generates the input pattern T to F_2 and activates a different STM pattern Y^* at F_2 since Y remains inhibited. If the top-down prototype due to Y^* also mismatches I at F_1, then the search for an appropriate F_2 code continues (adapted from Carpenter et al., 1991).

parameter can vary across learning trials a single ART system is able to encode widely differing degrees of generalization. Low vigilance leads to broad generalization and more abstract prototypes than high vigilance. In the limit of very high vigilance, prototype learning reduces to exemplar learning.With the help of the diagrams in figure 3, we will now follow in detail a typical ART search cycle. Not shown in this figure is the preprocessing field F_0 whose main purpose is a normalization of the input pattern.

(a) After the preprocessing by field F_0, an input pattern I generates a pattern of activity X at field F_1. The 2/3 rule is satisfied here because input I also activates the gain control at the F_1 level. The activation of the gain control is nonspecific because it does not depend on the type of pattern but only on its overall input activity. Pattern X both inhibits A and generates an output signal S from field F_1. Inhibition of A is necessary because otherwise a reset of field F_2 would occur. The signal S is multiplied by the bottom-up connection weights and results in a signal T that inputs to the F_2 level. The signal T produces an output Y from the level F_2. Here also the 2/3 rule is obeyed because the input signal I also nonspecifically activates the gain control for the F_2 level. The signal Y, for example, could result from the activation of the node(s) whose connection weights best matched the signal S.

(b) the pattern Y now generates a top-down signal pattern U which, after being multiplied by the top-down connection weights, results in the prototype pattern V. This prototype pattern V is compared at F_1 with the input pattern I. The result of this comparison is a new pattern of activity X* at F_1. If V mismatches I at F_1 the resulting activity X* will have significantly dropped. As a result of this reduction in total activity, less inhibition results at A.

(c) if now the vigilance criterion at A fails to be met, A can release a nonspecific signal to F_2 which inhibits the nodes at F_2 that were most active. As a result the signal Y is reset as well as the feedback signal U and its prototype V.

(d) pattern X is reinstated at F_1 and a different STM pattern Y* becomes active at F_2. If the top-down prototype due to Y* also mismatches I at F_1, then the search for an appropriate code continues until either a prototype has been found that satisfies the matching criterion at A, or a new category must be established at a previously uncommitted node.

In the sequel we will describe how the the ideas of this section can be implemented in the form of an algorithm for our Category ART. However, before we can explain the supervised Category ART algorithm, we first have to explain how a basic ART module works. As an example we take the Fuzzy ART module for unsupervised classification. This Fuzzy ART module will later be incorporated in the Category ART model.

3. Fuzzy ART algorithm

The Fuzzy ART component in Category ART consists of a preprocessing field of nodes, F_0, that modifies the current input vector **a** and a field F_1 that receives both bottom-up input from F_0 and top-down input from the field F_2. We do not need to distinguish between the connection weights of the top-down feedback paths and the bottom-up feedforward paths between the fields F_1 and F_2 in the Fuzzy ART module, both will be implemented by the same weights.

Three parameters determine the dynamics of a Fuzzy ART network, a *choice* parameter $\alpha > 0$; a *learning rate* parameter $\beta \in [0, 1]$ and a *vigilance* parameter $\rho \in [0, 1]$. The influence of these parameters on the network dynamics will be explained in the following paragraphs.

3.1. Preprocessing

When an M-dimensional input vector **a** is presented to the network it is first normalized by the field F_0. This normalization is necessary to guarantee stable category learning. The F_0 output activity vector **I** is a simple function of the F_0 input vector **a**, and its complement vector $\mathbf{a}^c$, namely, $\mathbf{I} = (\mathbf{a}, \mathbf{a}^c) = (a_1, ..., a_M, 1-a_1, ..., 1-a_M)$, where all a_i are in the interval [0, 1]. The net result of this normalization operation is accordingly a doubling of the length of the input vector **a**, while at the same time the norm of the new vector will always be equal to M. We use the following definition of the norm of a vector **x**

$$|\mathbf{x}| = \sum_{i=1}^{M} |x_i|.$$

We then get for the norm of **I**

$$|\mathbf{I}| = |(\mathbf{a}, \mathbf{a}^c)| = \sum_{i=1}^{M} a_i + \sum_{i=1}^{M} (1 - a_i) = M$$

3.2. Category choice

The input vector **I** is now fed forward from the F_1 field to the F_2 field. Both fields are implemented with a single layer of, respectively, $2M$ and N nodes. N is the capacity of the F_2 field and at the same time represents the maximum number of categories that this field can accommodate. All nodes in one layer are fully connected with all the nodes of the other layer, i.e., each of the N category nodes in the F_2 field has $2M$ connection with field F_1. Initially, before any learning has occurred all weights in the vectors $\mathbf{w}_j$ have the value 1 and each category node is said to be *uncommitted*. A weight vector $\mathbf{w}_j$ is also called a *template*.

When a pattern **I** is presented at field F_1, a choice function T_j is defined according to the following formula

$$T_j(\mathbf{I}) = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|},$$

where α is the choice parameter and $\wedge$ is the fuzzy AND operator, defined as

$$(\mathbf{x} \wedge \mathbf{y})_i = \min(x_i, y_i).$$

The fuzzy AND operator reduces to the Boolean AND operator in the case of binary vectors. The system is said to make a category choice when at most one F_2 node can become active at a given time. The F_2 node with maximum T_j will be chosen to represent the pattern **I**, and, when the Jth category node is chosen, the output vector **y** of the field F_2 is set as $y_J = 1$ and $y_j = 0$ if $j \neq J$. In a choice system, the F_1 activity vector **x** obeys the equation

$$\mathbf{x} = \begin{cases} \mathbf{I} & \text{if } F_2 \text{ is inactive} \\ \mathbf{I} \wedge \mathbf{w}_J & \text{if } J\text{th } F_2 \text{ node is chosen.} \end{cases}$$

If the chosen category J meets the vigilance criterion, that is if

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} \geq \rho,$$

then learning can occur. Mismatch reset occurs when the vigilance criterion is not met, and subsequently a new node is chosen. This search process continues until the chosen node satisfies the vigilance criterion. The search order among the nodes in the F_2 layer depends on the choice parameter α. If α is small then the search is more dominated by the pattern with the largest ratio $|\mathbf{I} \wedge \mathbf{w}_j| / |\mathbf{w}_j|$ than by the the size of $|\mathbf{I} \wedge \mathbf{w}_j|$ alone. For larger values of α we see that the patterns for which $|\mathbf{I} \wedge \mathbf{w}_j|$ is large dominate the search. We can now make the following hierarchy for the F_2 nodes that will be chosen when an input pattern $\mathbf{I}$ is presented at the F_1 layer (Huang et al., 1995):

(a) If there is a subset node it will be chosen over an uncommitted node. A subset node has a template $\mathbf{w}_j$ whose components satisfy

$$w_{ji} \leq I_i, \qquad i = 1..2M$$

This means that for a subset node

$$\frac{|\mathbf{I} \wedge \mathbf{w}_j|}{|\mathbf{w}_j|} = 1$$

(b) Because of the choice parameter $\alpha > 0$, among all the subset nodes the node with the largest template $\mathbf{w}_j$ will be chosen first.

(c) An uncommitted node will be chosen whenever there are no subset nodes and all committed nodes j satisfy

$$\frac{|\mathbf{I} \wedge \mathbf{w}_j|}{|\mathbf{w}_j|} \leq \frac{1}{2}$$

In our implementation of the Fuzzy ART algorithm we have changed this biologically oriented blind search. Mainly for reasons of efficiency, we always maintain a list of committed and uncomitted nodes to speed up the search process.

3.3 Learning

The template vector $\mathbf{w}_j$ is updated according to the following equation

$$\mathbf{w}_J^{(new)} = \beta(\mathbf{I} \wedge \mathbf{w}_J^{(old)}) + (1-\beta)\mathbf{w}_J^{(old)}$$

When $\beta = 1$ we speak of *fast* learning. For efficient coding of noisy inputs, we choose the fast learning option when J is an uncommitted node, and then take $\beta < 1$ after the node is committed. Then $\mathbf{w}_J^{(new)} = \mathbf{I}$ the first time category J becomes active. After the commitment the weight vector update causes the new weight vector to become more aligned with the most recently coded input pattern.

4. Category ART algorithm

Our Category ART neural network module is a simplification of the ARTMAP module. In the ARTMAP architecture, as shown in figure 1, two ART modules,

ART_a and ART_b, are linked together via an inter-ART module, F_{ab}, called the map field. The ART_a and ART_b, modules could be one of the set ART1, ART2-A or Fuzzy ART. Two choices are described in the literature, in ARTMAP (Carpenter et al., 1991a) two ART1 modules are combined and in Fuzzy ARTMAP (Carpenter et al., 1992) two Fuzzy ART modules are combined. Input vectors to ART_a and ART_b are named **a** and **b**, respectively, and $\mathbf{x}^a$ and $\mathbf{x}^b$ are the outputs of the corresponding F_1 fields, F_{1a} and F_{1b}, and $\mathbf{y}^a$ and $\mathbf{y}^b$ are the outputs of the corresponding F_2 fields, F_{2a} and F_{2b}. For the map field, let $\mathbf{x}^m$ denote its output vector and $\mathbf{w}_{mj}$ denote the weight vector from the j-th node F_{2a} to F_m. The map field includes an associative memory and control signals and both are used to form predictive associations between categories of ART_a and ART_b and to realize the match tracking rule. Match tracking means that a wrong prediction triggers the search mechanism in the ART_a module anew to look for a better match or, if a better match cannot be found, for a new category. Match tracking can reorganize category structure so that predictive errors will not be repeated on subsequent presentations of the same input. ARTMAP can be used for mapping multidimensional vectors. However, when we want to associate category labels with multidimensional vectors, for example vowel labels with spectral representations, using ARTMAP forces us to represent the category labels as a multidimensional input vector to the ART_b network, and initialising ART_b's vigilance to a very high level. Two possible options how to choose ART_b's input vector when we have M different categories are: choose vector **b** of dimension M and make $b_i = 1$ for category i, or use a binary representation with a p-dimensional vector **b**, where $2^p >= M$. This vector **b** is processed by the ART_b module, in which different input vectors (resulting from different category labels) should activate different output nodes of field F_{2a}. In effect, ART_b categorizes one to one, each different input is represented by a different output node. This means that for each **b**, only one output node is active and thus the norm of $\mathbf{y}^b$ equals one. As a consequence this makes the map vigilance parameter, ρ_{ab}, in the following match tracking equation equation, which is equation (35) in the Fuzzy ARTMAP implementation of Carpenter et al. (1992), non-effective.

$$\left|\mathbf{x}^{ab}\right| < \rho_{ab}\left|\mathbf{y}^{a}\right|$$

We note that both in the ARTMAP as well as in the Fuzzy ARTMAP implementations of Carpenter at al. (1991a, 1992) the map vigilance parameter is ineffective because the output of the ART_b network, $\mathbf{y}^b$, is always normalized to one.

In our Category ART algorithm, the second ART system, ART_b, whose only function is to form a category representation, and the map field are replaced by an ordered collection of category labels and an array of pointers. There is a pointer to a category label for each node of the F_2 layer of the ART_a module.

The Category ART learning algorithm in pseudo code goes as follows:

```
for all (pattern p, categoryLabel c)
    learn (pattern p, categoryLabel c)
          if categoryLabel not in categoryLabelList
                create categoryLabel c
                add categoryLabel c to categoryLabelList
          endif
          J = categorize p by ARTa network
          if categoryLabelList[nodePointer[J]] ≠ c
                temporarily increase vigilance
                J = categorize p by ARTa network
```

```
            reset vigilance
        endif
        update_weights (w_J)
        nodePointer[J] = index categoryLabel in categoryLabelList
    end learn
endfor
```

Because of the combination of match tracking and fast learning, a single ARTMAP system can learn a prediction for a rare event that is different from that for a cloud of similar frequent events in which it is embedded. This means that eventually noise is also learned since the system cannot know beforehand what constitutes the signal and what the noise.

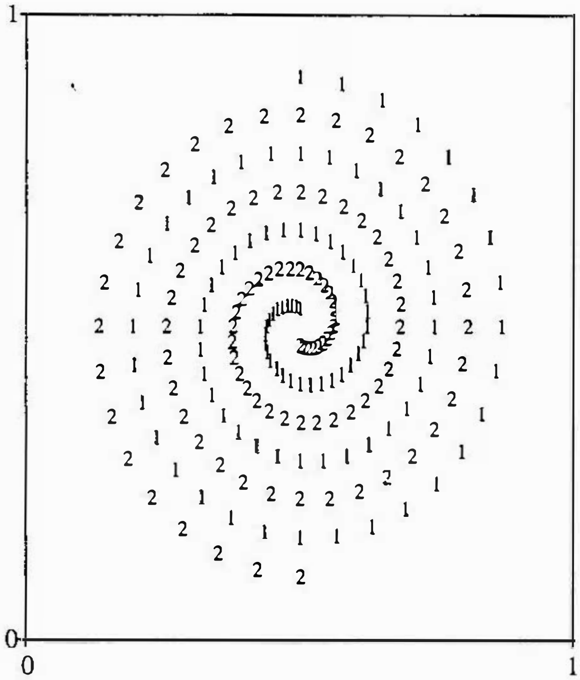

Figure 4. Two spirals in the plane. Each spiral consists of 97 points.

5. Simulation: Learning to tell two spirals apart

To get an impression of the possibilities of a Category ART network we will describe its performance on a complicated classification task. We reproduce the example from Carpenter et al. (1992) in which they describe the Fuzzy ARTMAP network: learning to tell two spirals apart in a two dimensional plane. This benchmark task cannot be learned by a standard feedforward network with backpropagation. According to the authors cited above, Lang and Witbrock (1988) succeeded by constructing a special 2-5-5-5-1 network with each node connected to all nodes in subsequent layers. The system had 138 trainable weights. With their fastest algorithm they needed at least 8000 epochs to complete the task, i.e., each of the 194 points in the training set responds to within 0.4 of its target output value. An epoch is one full presentation of the entire training set.

The spirals of the benchmark each make three complete turns in the plane and consist of 97 points as is shown by figure 4 above. The coordinates of the points of the two spirals are

$$x_n^{(1)} = 1 - x_n^{(2)} = r_n \sin\alpha_n + 0.5$$
$$y_n^{(1)} = 1 - y_n^{(2)} = r_n \cos\alpha_n + 0.5$$

where

$$r_n = 0.4(\frac{105-n}{104})$$

and

$$\alpha_n = \frac{\pi(n-1)}{16}.$$

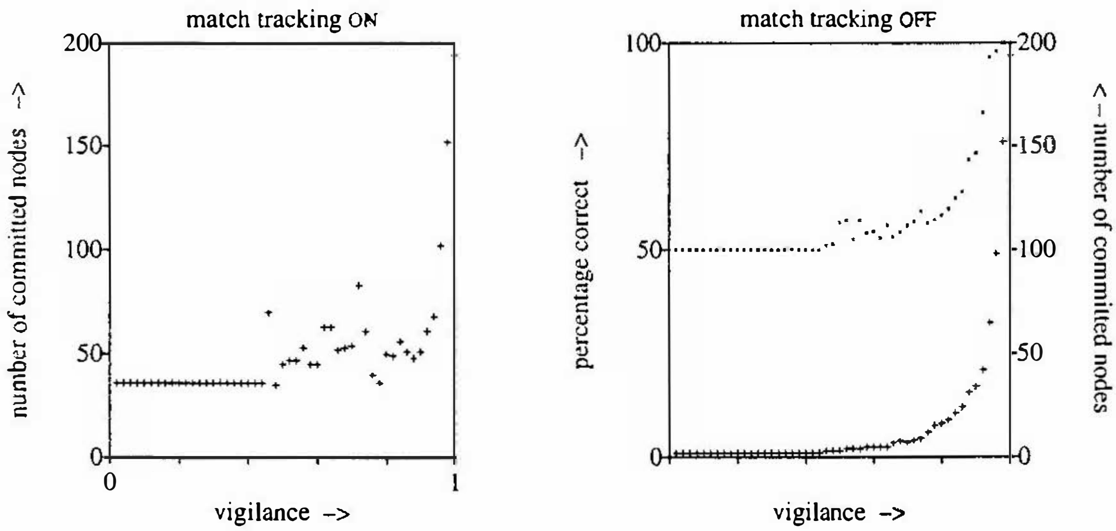

Figure 5. Fuzzy Category ART performance figures for the two spirals data set. The figure on the left shows the number of committed nodes as a function of the vigilance parameter when match tracking was active. In this case there was always 100% correct classification. For match tracking off, the figure on the right shows besides the number of committed nodes (+) also the percentage correct classification (x) as a function of the vigilance parameter.

A trivial solution with Category ART is obtained by selecting for the vigilance parameter $\rho = 1$. In this case the network learns all patterns in one epoch with 100% correct classification. However, the network uses 194 category nodes for the classification, one node for each training pattern. This amounts to using 770 parameters for the classification: the 194 times 4 connection weights from the F_2-nodes in the Fuzzy ART module plus 194 category index pointers to either the first or the second spiral. In a Fuzzy Category ART we have in principle four parameters that determine learning. The first three parameters are determined by the Fuzzy ART module namely the choice parameter α, the vigilance parameter ρ, and the learning parameter β. The fourth parameter *matchtrack* determines whether matchtracking is on or off. The parameters that influence most the number of categories and therefore the number of weights, are the vigilance parameter and the matchtrack parameter. When matchtracking is on, the network is capable of raising its vigilance level when a mismatch at the category index level occurs. The most effective strategy to lower the number of categories is to start with matchtracking on and a very low vigilance, i.e.,

$\rho = 0$. We performed two series of runs with the vigilance parameter increasing from 0 in steps of size 0.02 to 1.0, the learning parameter β fixed at 1, the choice parameter α fixed at 0.001. The first series had match tracking on, the second had match tracking off. The results are displayed in figure 5. For all combinations of the parameters the training of Category ART completed within 15 training epochs. When match tracking was on, the percentage correct classification obtained was always 100%. The left plot in figure 5 shows the number of committed nodes as a function of the vigilance level. For vigilance levels smaller than approximately 0.45 the number of committed nodes stays at the very low value of 36. It shows a gradual increase in the number of committed nodes when the vigilance level increases to 0.96, still higher values of the vigilance level show a steep increase in this number. The maximum, 194, is reached when the vigilance level is equal to 1.0. When match tracking is off, the percentage correct drops to 50% when the vigilance level is reduced, as the right plot in figure 5 shows. Because match tracking is off, the number of committed nodes drops much steeper, ultimately to only two committed nodes when the vigilance level drops below 0.4.

5. Conclusions

The preliminary performance of the Category ART neural network is satisfactory. It ia capable of learning reasonably complex tasks in a very short time. The simulation study in which the network task was to tell two spirals apart was handled well. This task was already too complex for a standard feedforward network with backpropagation (BP). As Lang and Witbrock (1988) showed, only a specially constructed BP network could handle this task after an enormous amount of training epochs (8000). Category ART was able to handle the task several orders of magnitude faster, i.e., within 15 epochs. This neural network design therefore shows great potential for vowel recognition tasks, a complex task, with interaction between spectral cues and fundamental frequency. Further classification tests with these type of networks will be performed on vowel data from the TIMIT corpus (Lamel et al., 1986).

6. Acknowledgement

The author wants to thank Louis Pols for his critical and constructive comments during this study.

7. 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., Grossberg, S. and Reynolds, J.H. (1991a), "ARTMAP: Supervised real-time learning and classification of non-stationary data by a self-organizing neural network", *Neural Networks* **4**, 565–588.

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.

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**, 698–712.

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

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

Grossberg, S. (**1986**), "The adaptive self-organization of serial order in behavior: speech, language and motor control", in *Pattern Recontion By Humans And Machines, Volume I: Speech Perception*, E. Schwab & H. Nusbaum (eds.), Academic Press, Inc.

Huang, J., Georgiopoulos, M. and Heileman, G.L. (**1995**), "Fuzzy ART properties", Neural Networks 8, 203–213.

Lamel, L.F., Kassel, R.H., and Seneff, S. (**1986**), "Speech Database Development: Design and Analysis of the Acoustic-Phonetic Corpus", *Proc. DARPA Speech Recognition Workshop*, Report NO. SAIC-86/1546, 100–109.

Lang, K.J. and Witbrock, M.J. (**1988**), "Learning to tell two spirals apart", in *Proc. 1988 Connectionist Models Summer School*, 52–59.

Weenink, D.J.M. (**1992**), "Introduction to neural nets", *Proceedings of the Institute of Phonetic Sciences University of Amsterdam* **15**, 1–25.