To appear in Proceedings of the European Conference on Computational Intelligence, Košice, Slovakia, Aug 30 - Sep 1, 2000. Eds. Sinčák, P., Vaščák J., Kvasnička, V., Mesiar, R. Published in the series Advances in Soft Computing, Springer-Verlag, Berlin, Sep 2000.

Graded Signal Functions for ARTMAP Neural Networks

Norbert Kopčo and Gail A. Carpenter

Department of Cognitive and Neural Systems, Boston University
677 Beacon St., Boston, MA 02215, USA
{kopco, gail}@cns.bu.edu

Abstract. *This study presents an analysis of a modified ARTMAP neural network in which a graded signal function replaces the standard choice-by-difference function. The modifications are introduced mathematically and the performance of the system is studied on two benchmark examples. It is shown that the modified ARTMAP system achieves classification accuracy superior to that of standard ARTMAP, while retaining comparable complexity of the internal code.*

Keywords*. ARTMAP, fast learning, graded signal function, neural network*

1. Introduction

Adaptive Resonance Theory (ART) was introduced by Grossberg [1] as a theory of human cognitive information processing. Based on the theory, a series of real-time neural network architectures for unsupervised and supervised learning have been developed. These networks combine fast learning with stable category coding and are a suitable tool for many pattern recognition problems. The ART models for unsupervised learning include ART 1, ART 2, fuzzy ART, and distributed ART. ARTMAP, a family of supervised ART architectures developed for classification problems, includes the fuzzy ARTMAP and distributed ARTMAP neural networks. A collection of papers on ART models can be found in [2], and more recent models are summarized in [3].

The present paper focuses on the process of search for a best category code in response to a given input in ARTMAP networks. Specifically, a new signal function is proposed that enables the system to find near-optimum discrimination curves between categories in complex input space. The modified signal function is introduced mathematically, then evaluated by implementing it in a fuzzy ARTMAP system and analyzing its performance on two benchmark problems.

2. Description of Fuzzy ARTMAP Dynamics

This section gives a brief summary of the fuzzy ARTMAP [4] algorithm. The **inputs** of the fuzzy ARTMAP system are usually normalized by complement coding, which converts an M-dimensional input vector $\boldsymbol{a} = (a_1,\dots,a_M)$ $(0 \le a_i \le 1)$ into $2M$-dimensional input pattern $\boldsymbol{I} = (\boldsymbol{a}, 1-\boldsymbol{a}) = (a_1,\dots,a_M, 1-a_1,\dots,1-a_M)$. The pattern is normalized since $|\boldsymbol{I}| = M$, where $|\boldsymbol{I}| \equiv \sum_{i=1}^{2M} |I_i|$ is the city-block norm.

When a new input is presented, the system **searches** for a candidate coding node within its coding layer. In ART systems, the j-th coding node defines a hyper-rectangle R_j, or *coding box*, in the M-dimensional input space, described by the weights $w_{i,j}$ leading to that node. The hyper-rectangle reduces to a rectangle in two dimensions or to an interval in one dimension (Figure 1). For every input pattern, the ARTMAP search mechanism chooses the smallest coding box that is covering the input, or the box that is closest to the input, based on the activation of the *choice-by-difference* (CBD) signal function T_j [5], now used in a majority of simulations. The CBD signal function is defined as:

$$T_j = M(2-\alpha) - d(R_j, \boldsymbol{a}) - \alpha|R_j| = |\boldsymbol{w}_j \wedge \boldsymbol{I}| + (1-\alpha)\left(2M - |\boldsymbol{w}_j|\right). \quad (1)$$

In this equation, α is a parameter (usually $\alpha = 0^+$), $d(R_j, \boldsymbol{a})$ represents the city-block distance from the input pattern $\boldsymbol{a}$ to the coding box R_j, and $|R_j|$ represents the size of R_j. The J-th coding node is chosen as a candidate code if its signal function T_J has the maximum value.

The candidate node is then compared with the input pattern according to a **match rule**. The candidate **resonates** if $|\boldsymbol{I} \wedge \boldsymbol{w}_J| > \rho \boldsymbol{I}$; or it is **reset** if the inequality does not hold, where $\rho \in [0,1]$ is called a *vigilance* parameter. If reset occurs, a search for a new candidate is initiated, or a new coding node is created. If the candidate node resonates, the system checks whether the node is associated with the correct output class (always satisfied for new nodes). If the node is associated with an incorrect category, a process of **match tracking** is initiated, i.e., ρ is increased just enough so that the

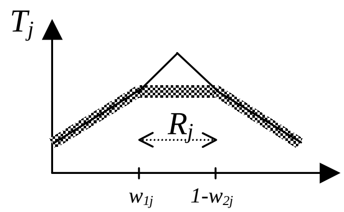

Figure 1 Choice signal for standard CBD (▩) vs. graded CBD (—) in one input dimension

current candidate will not resonate any more and a search for a new candidate is initiated.

Once a coding node is found that satisfies all the requirements, **learning** is initiated that updates all the weights leading to the *J*-th node, defined by

$$\boldsymbol{w}_J^{(new)} = \beta\left(\boldsymbol{I} \wedge \boldsymbol{w}_J^{(old)}\right) + (1-\beta)\boldsymbol{w}_J^{(old)}. \qquad (2)$$

Fast learning is usually chosen, obtained by setting $\beta=1$.

3. Definition of Graded Signal Functions

In general, the ARTMAP search for the internal code that best matches the presented input pattern can be accomplished by choosing one of many different signal functions, used to determine the activation of the coding nodes. Most current ARTMAP systems use the *choice-by-difference* signal function (CBD, [5]), which implements the idea of minimum fast learning. For the CBD function the signal is independent of the position of the input pattern if the input is located within the coding box, as shown for one dimension in Figure 1. In the present paper, a new signal function is introduced, called a *graded choice-by-difference* function, or graded CBD, which makes the choice signal dependent on the input position even when the input lies within the category box. Namely, an input near the center of the box R_j generates a larger signal T_j than an input near the boundary of the box (Figure 1).

The activation in the graded CBD signal function is defined by:

$$T_j = M(2-\alpha) - d(R_j, \boldsymbol{a}) - \alpha\left|R_j\right|(1-\eta\gamma_j) \qquad (3)$$

where η is a parameter that defines by how much the activation at the center of the box is increased relative to the box boundaries. When η=0, graded CBD reduces to standard CBD (1). In (3), γ_j specifies the minimum of graded activations across dimensions $i=1\ldots M$:

$$\gamma_j = \min_{i=1..M}\left[1-\left[\frac{2\left|a_i - c_{j,i}\right|}{1-w_{j,i+M} - w_{j,i}}\right]^+\right]^+ \tag{4}$$

In (4), $c_{j,i} \equiv (1-w_{j,i+M} + w_{j,i})/2$ denotes the center of the j-th coding box in the i-th dimension, and $[a]^+ \equiv \max(a,0)$ is a rectification operator. Note that $\gamma_j = 1$ at the center of R_j and $\gamma_j = 0$ at any point $\boldsymbol{a}$ on the boundary of R_j. In order to ensure that the same input would choose the same category if it were immediately re-presented (direct access), the ART match rule was also modified, to better correspond to the new choice rule. In addition to the match criterion defined above, the new match rule essentially simulates the process of weights-update (2) followed by re-presentation of the current input. Then, the J-th node resonates only if the simulated update led to the desired choice of the winning node by direct access. The resulting graded signal function system has the capacity to create more accurate decision boundaries, especially when these boundaries are not parallel to the input space axes (Figure 2).

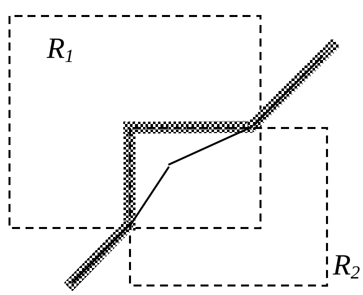

Figure 2 Decision boundaries between two category boxes (R_1 and R_2) with standard CBD (▩) vs. graded CBD (—)

4. Results on Benchmark Data and Discussion

The performance of a fuzzy ARTMAP system with the graded CBD signal rule was evaluated on two benchmark problems, the circle-in-the-square (CIS) problem and a diagonal problem. Data sets for both problems consist of 2-dimensional uniformly distributed points, with the values in each dimension ranging from 0 to 1. Each data set has two output classes. In CIS, a point $\boldsymbol{a}=(a_1,a_2)$ is in the class C_{out} if $(a_1-0.5)^2+(a_2-0.5)^2>\frac{1}{2\pi}$, otherwise it is from the class C_{in}. In the diagonal data set a point is from the class C_{lower} if $a_1>a_2$, otherwise it is from the class C_{upper}. Simulations with training sets of different sizes (100, 1000, or 10,000 points) were performed. The testing set size was fixed to 10,000 points.

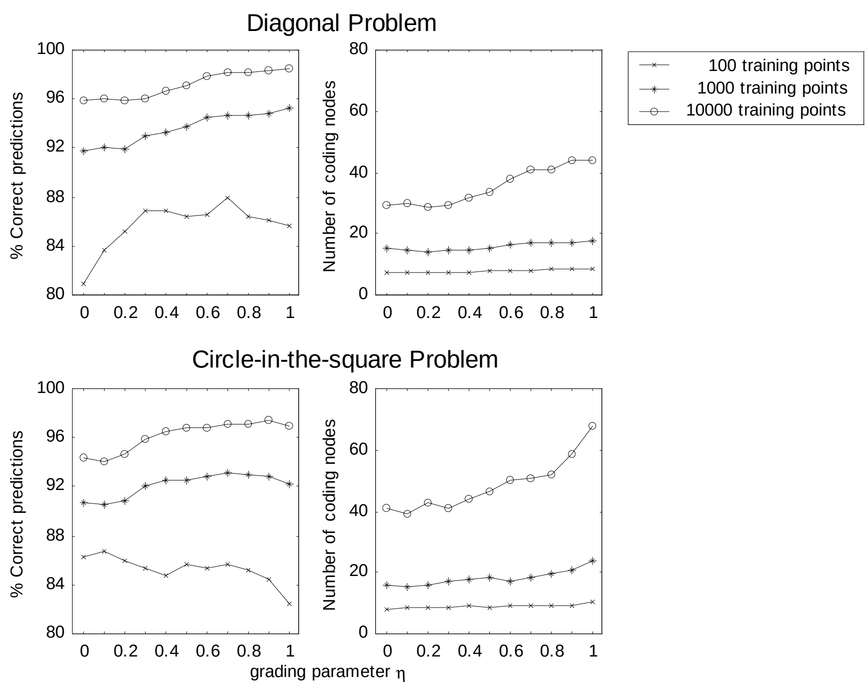

Figure 3 Simulations of fuzzy ARTMAP with standard CBD (η=0) and with graded CBD signal function (η>0). The upper row shows results of simulations with the diagonal data set, the lower row contains data for circle-in-the-square simulations

The results of simulations of the two benchmark problems are shown in Figure 3, which shows percent correct predictions and number of coding nodes as functions of the value of the graded signal parameter η. In the graphs, η=0 corresponds to the standard CBD rule. Each point in Figure 3 corresponds to an average of 10 simulations with randomized order of inputs. For each of the conditions, application of the graded signal function led to improved performance, accompanied in some conditions by a slight increase in the internal code complexity. This improvement is mainly due to the improved ability of the new signal function to approximate decision boundaries not parallel to the axes of input feature space (Figure 2).

These results indicate that ARTMAP systems with the graded CBD signal rule can be used for many types of pattern recognition problems, especially when the data from individual classes are not easily separable, which may lead to many overlapping category boxes. More simulations are necessary, especially with noisy data, to better understand the behavior of the system in complex environments.

[This research was supported by grants from the Office of Naval Research and the Defense Advanced Research Projects Agency (ONR N00014-95-1-0409 and ONR N00014-1-95-0657).]

References

1. Grossberg, S. (1976) Adaptive pattern recognition and universal recoding, II. Feedback, expectation, olfaction, and illusions. Biological Cybernetics. **23**, 187-202

2. Carpenter, G.A., Grossberg, S. (1991) Pattern Recognition by Self-Organizing Neural Networks. M. I. T. Press, Cambridge Massachusetts

3. Carpenter, G.A., Milenova, B.L., Noeske, B.W. (1998) Distributed ARTMAP: a neural network for fast distributed supervised learning. Neural Networks. **11**, 793-813

4. 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**: 698-713.

5. Carpenter, G.A., Gjaja M.N. (1994) Fuzzy ART choice functions. In Proceedings of The World Congress on Neural Networks (WCNN). **4**, 133-142