FUZZY LAPART SUPERVISED LEARNING
THROUGH INFERENCING FOR STABLE CATEGORY RECOGNITION

Gabsoo Han, Fredric M. Ham, and *Laurene V. Fausett
Florida Institute of Technology
Department of Electrical and Computer Engineering
*Department of Applied Mathematics
150 West University Boulevard
Melbourne, Florida 32901-6988

Abstract — Fuzzy LAPART (laterally primed adaptive resonance theory), a neural network architecture for supervised learning through logical inferencing, is introduced with fast and slow learning algorithms and match tracking capability. Based on the original architecture developed by Healy, et al., the enhanced architecture consists of interconnected fuzzy adaptive resonance theory (fuzzy ART) modules originated by Carpenter, et al. The interconnections enable fuzzy LAPART to infer one pattern class from another to form a predictive pattern class. Slow learning capability has been incorporated into the neural network with fast commit and slow recode options. The problem of separation of spirals is used to perform benchmark tests for fuzzy LAPART. Also, based on fuzzy set theory, geometric interpretations are presented in 2 and 3 dimensional spaces using fuzzy LAPART. Performance results for both test cases are compared to results obtained from a counterpropagation clustering network. Fuzzy LAPART shows outstanding performance in the benchmark tests, with more efficient ability of clustering than counterpropagation. Slow learning (fast-commit and slow-recode) was used for fuzzy LAPART for 3-dimensional category space interpretations.

overcomes this problem. With the ART1 modules in LAPART replaced with two fuzzy ART systems, fuzzy LAPART can be used for more general classes of problems. In addition to the incorporation of the coupled fuzzy ART modules, match tracking capability has been included in the fuzzy LAPART architecture, which increases the ART vigilance parameter by the minimum amount necessary to correct a predictive error. Thus, fuzzy LAPART can realize a minimax learning rule that conjointly minimizes predictive error and maximizes code compression, or generalization. Slow learning with fast-commit and slow-recode options have also been incorporated into the neural network architecture. For efficient coding of noisy input sets, it is useful to use fast commitment and slow recoding training capability. The problem of separation of spirals is used to perform benchmark tests for fuzzy LAPART. Also, based on fuzzy set theory, geometric interpretations are presented in 2 and 3 dimensional spaces using fuzzy LAPART. Performance results for both test cases are compared to results obtained from a counterpropagation clustering network. Two types of counterpropagation networks are compared. The only difference in the two types is that one utilizes a topological update and the other does not.

I. INTRODUCTION

ART (adaptive resonance theory), developed by Carpenter and Grossberg [1], has gained much attention for solving the stability-plasticity dilemma. Successively, ART2 [2] and fuzzy ART [3] were developed to address the problem of handling both binary and analog inputs. Recently, Carpenter, et al. [4] developed a new supervised coupled fuzzy ART network, called fuzzy ARTMAP, which is capable of self-organizing recognition categories to arbitrary sequences of the input patterns. More recently, Healy, et al. [5] developed the new architecture, LAPART (laterally primed adaptive resonance theory), for logical inferencing and supervised learning, which is easier to implement than fuzzy ARTMAP [4]. In addition, LAPART can be used to recognize sequences of input data. However, LAPART has been restricted to binary inputs because of the two coupled ART1 modules. In this work, a new LAPART architecture has been developed, fuzzy LAPART, which

II. FUZZY SET THEORY AND FUZZY ART

The fuzzy ART system performs computations based on fuzzy set theory [6] using the ART1 architecture. The fuzzy ART model was developed by Carpenter, et al. [3] by generalizing ART1 such that it is capable of learning stable recognition categories in response to both analog and binary input patterns. For the special case of binary inputs and fast learning, the computations of fuzzy ART are identical to those of the ART1 neural network. The learning algorithm for fuzzy ART [7] can be described as follows. The weight vector w_J is updated according to the equation

$$w_J^{(new)} = \beta(I \wedge w_J^{(old)}) + (1 - \beta)w_J^{(old)} \quad (1)$$

Fast learning corresponds to setting $\beta = 1$. Complement coding establishes a connection between on-cell/off-cell representations and fuzzy set theory

operations. For definiteness, let the input set consist of two-dimensional vectors **a** preprocessed into four-dimensional complement coding form [6]. Thus,

$$I = (a, a^c) = (a_1, a_2, 1 - a_1, 1 - a_2) \qquad (2)$$

In this case , each category j has a geometric representation as a rectangle R_j, as follows. Following Eq. 2, the weight vector W_j can be written in complement coding form as

$$w_j = (u_j, v_j^c) \qquad (3)$$

where u_j and v_j are two-dimensional vectors. Let the vector u_j define one corner of a rectangle R_j and let v_j define another corner of R_j (see Fig. 1 (a)). The size of R_j is given by

$$|R_j| \equiv \sum_{i=1}^{N} [\max_j\{a_i\} - \min_j\{a_i\}]$$
$$= [\ |\vee_j a| - |\wedge_j a|\ ] \qquad (4)$$

Learning increases the size of each R_j, in fact, the size of R_j grows as the size of w_j shrinks during learning, and the maximum size of R_j is determined by the vigilance parameter ρ. During each fast-learning trial, R_j expands to $R_j \oplus$ **a** (see Fig. 1 (b)). The corners of $R_j \oplus$ **a** are given by **a** $\wedge u_j$ and **a** $\vee v_j$, where

$$(x \vee y)_i \equiv \max(x_i, y_i) \qquad (5)$$

Hence, by Eq. 4, the size of $R_j \oplus$ **a** is given by

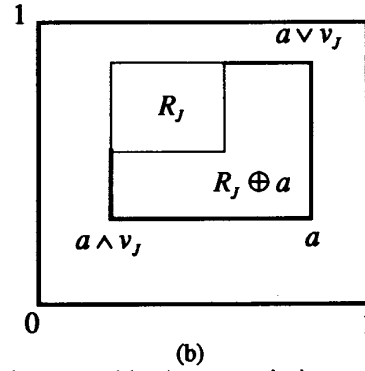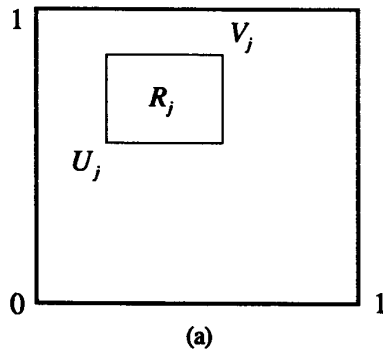$$|R_j \oplus a| = |(a \vee v_j) - (a \wedge v_j)| \qquad (6)$$

(a)

(b)

Figure 1. (a) A geometric interpretation as a rectangle R_j, (b) Fast learning with expanding R_j to $R_j \oplus a$.

A geometric interpretation in 3 dimensions of fuzzy ART with fast-commit slow-recode can now be illustrated. Fuzzy ART was trained with 8 input patterns (shown in Table I), 4 cluster units were assigned to the network, and a vigilance parameter of 0.8 was used. After 34 iterations, the weights of the network converged. Figure 2 illustrates the fast-commit and slow-recode procedures. With β=0.5 in Eq. 1 , the network training is designed to stop when the change in the weights remains within 10^{-10}.

TABLE I
3-D Training Patterns for Fuzzy ART

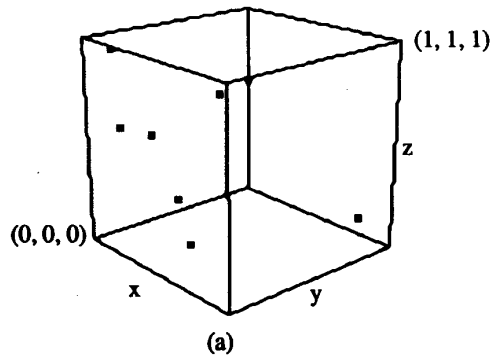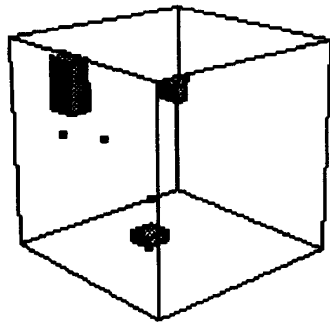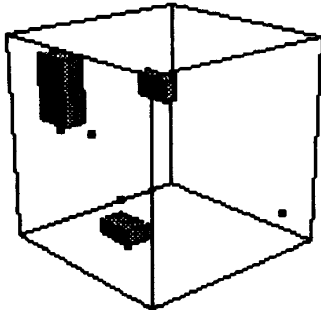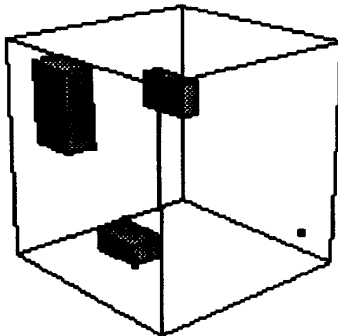|	
pattern 1 =	{0.13 0.24 0.76}
pattern 2 =	{0.42 0.35 0.2}
pattern 3 =	{0.33 0.26 0.54}
pattern 4 =	{0.29 0.16 0.57}
pattern 5 =	{0.73 0.86 0.24}
pattern 6 =	{0.66 0.29 0.71}
pattern 7 =	{0.19 0.47 0.26}
pattern 8 =	{0.86 0.24 0.79}

(a)

Figure 2. Training patterns in 3 dimensional space.

(b)

(c)

(d)

Figure 2. (*continued*) Training patterns in 3 dimensional space. (a) training patterns, (b) after 2 iterations, (c) after 10 iterations, (d) after 34 iterations.

III. FUZZY LAPART

The LAPART architecture includes two ART modules (ART_a and ART_b,) that create stable recognition categories in response to sequences of

input patterns. Binary LAPART employs ART1 systems for the ART_a and ART_b modules, whereas fuzzy LAPART substitutes fuzzy ART systems for these modules (see Fig. 3). During supervised learning, ART_a receives a stream $\{a^{(p)}\}$ of input patterns, and ART_b receives a stream $\{b^{(p)}\}$ of input patterns, where $b^{(p)}$ is the correct prediction given $a^{(p)}$. The basis for the recognition function of fuzzy LAPART is the coupling of two pattern classifier networks through a system of interconnects. In the simplest version of fuzzy LAPART, the two classifier networks are fuzzy ART systems. The interconnects implement a dual system of inference rules. Recognition of a member of a familiar pattern class by one net A triggers a rule inferring that a member of a familiar class will be recognized by net B. The second rule enables the other network to veto the prediction if it is not consistent with available data causing *match tracking* in net A. Fuzzy LAPART learns the inferences during presentation of training pairs of patterns, through 1) fuzzy ART pattern classification, involving synaptic learning within each network, and 2) synaptic learning of the class-to-class inferences through interconnects (see Fig. 3).
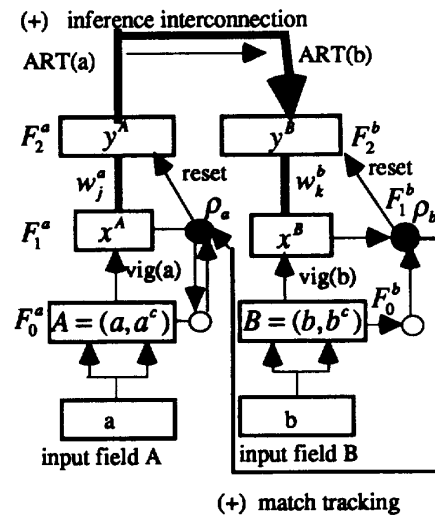

(+) inference interconnection

(+) match tracking
Figure 3. Fuzzy LAPART architecture.

IV. SIMULATION RESULTS: TWO SPIRAL PROBLEM

Training a neural network to distinguish two spirals is a benchmark task proposed by Lang and Witbrock [8]. The main purpose of this simulation is to compare the benchmark test performance results for fuzzy LAPART to a counterpropagation [9] clustering network, using the same training and test sets. Two

48

different counterpropagation networks are compared, one utilizes a topological update [10] and the other does not. Both counterpropagation networks were trained after 1000 epochs. The total number of FLOPs for the network without topological update was 2,328,002, the network with topological update required 11,640,002 FLOPs. Fuzzy LAPART was trained with three different baseline vigilance parameters, i.e., 0.9, 0.95 and 1.0. With the baseline vigilance parameter set to 0.9 and 0.95, fuzzy LAPART trained after 5 epochs, and the required number of FLOPs for the two different baseline vigilance parameters was 49,845 and 79,543, respectively. When the baseline vigilance parameter was set to 1.0, only one training epoch was required and 38,412 FLOPs. The fuzzy LAPART simulation results are depicted in Fig. 4, where the black dots indicate the category regions for a two-spiral training pattern with the baseline vigilance parameter $\bar{\rho}_a = 1.0$. Since $\bar{\rho}_a = 1.0$, the points in the figure also represent each training pattern. Since only 1 epoch is required to train the network, each input established its own category, thus creating a nearest neighbor classifier that is 100% correct for the 194 point spiral training set.
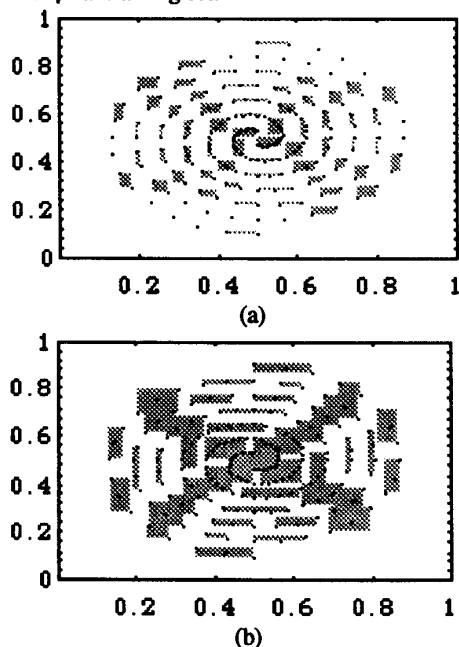

Figure 4. Black dots represent clusters with vigilance 1.0. (a) clusters with vigilance 0.95.(b) clusters with vigilance 0.9.

Moreover, 100% correct prediction was achieved using a test set consisting of a double dense spiral, each with 385 points. However, the necessary number of stored weights for fuzzy LAPART is 776,

two for each point on the two spirals. In order to reduce the number of necessary weights, the baseline vigilance parameter $\bar{\rho}_a$ is set to 0.95. This reduced the number of ARTa categories from 194 to 78 shown in Fig. 4 (a), which results in a 40% reduction Moreover, for the double dense spiral test case, 99.0% correct prediction was achieved. However, reducing $\bar{\rho}_a$ to 0.9 (see Fig. 4 (b)) generally resulted in an increase of the clustering region area (the maximum size was 0.2). The clustering region area can be calculated according to Eq. 6. Reducing $\bar{\rho}_a$ to 0.9 resulted in a reduction of ARTa nodes to 47. Now, the double dense spiral test case results showed 91.4% correct prediction, as indicated in Fig. 6. All fuzzy LAPART simulation results are shown in Fig. 6.
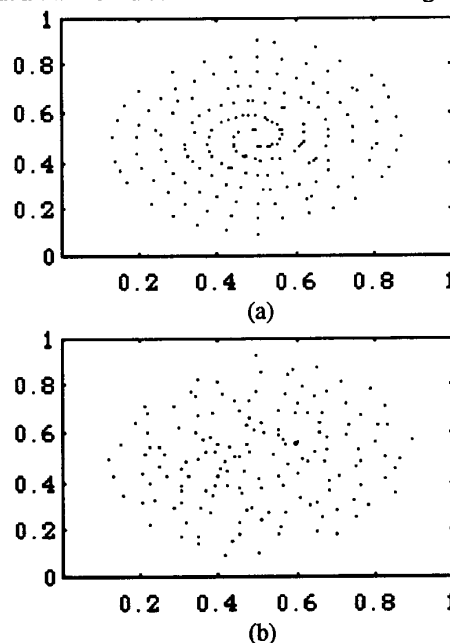

Figure 5. (a) Weight distributions between input layer and Kohonen layer after training using topological update, (b) weight distributions without topological update.

For the double dense spiral test simulations using counterpropagation, 88.4% correct prediction was achieved for the topological update network (the mean square error of correct prediction for this test case was 0.36231), and 5% correct prediction was achieved for the network without topological update (the mean square error of correct prediction for this test case was 1.33793). The results obtained using topological update can be seen in Fig. 5 (a), and the results without topological update are shown in Fig. 5 (b). The counterpropagation network that was trained with topological update, required a topological structure in

49

the Kohonen layer with neighborhood radii. As previously mentioned, a total of 1000 training epochs was required when the neighborhood radii were two. The simulations were performed with decreasing the neighborhood radii from R=2 to R=0 (neighborhoods of radii without "wrap around" was gradually decreased as the clustering process progressed with a linear decrease of the learning rate α). When simulations were run with the initial radii wider than two, it was observed that the network performance degraded relative to the network with initial radii set at 2. All initial weights were initialized with *uniform* distributions, by setting all weights to values between 0.1 and 0.9. As the results show in Fig. 6, using topological learning in the Kohonen layer gives remarkable improvement in performance. However, the overall performance of the counterpropagation network is still below 90%, which is less than that achieved with fuzzy LAPART. Results show that the performance of the network is heavily dependent on the ordering of the input training patterns and initialization of the weights. During training, input patterns were presented in random order. When the input patterns were presented sequentially for counterpropagation, as in the case of fuzzy LAPART, the network formed too few cluster units resulting in poor performance.

Figure 6. Test results for single and double dense spiral simulations.
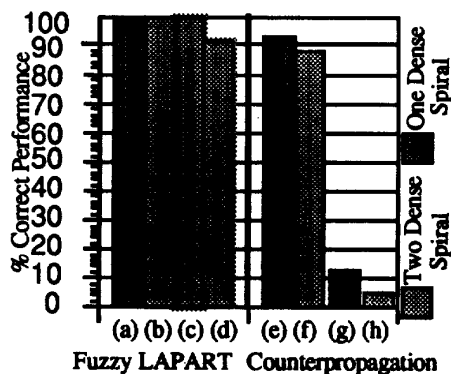
(a) test results with training patterns for all vigilance parameter cases (100%).
(b),(c), and (d) test results with test patterns for the baseline vigilance parameters 1.0, 0.95, and 0.9, respectively.
(e),(f) test results with topological update (88.4%).
(g),(h) test results without topological update.

V. CONCLUSIONS

Fuzzy LAPART, based on the original version of LAPART (laterally primed adaptive resonance theory) for supervised learning of recognition categories and multidimensional maps, has been introduced. The incorporation of fuzzy ART modules in the architecture allows for analog as well as binary inputs to the neural network. In addition, a match-tracking mechanism has also been incorporated in the architecture, which is based on the coupled fuzzy ART modules. With the match tracking capability, fuzzy LAPART functions similar to fuzzy ARTMAP, however, there are significant differences between the two networks due to the mapping fields and match tracking channels of each architecture. In fuzzy LAPART, the intermediate layer (the so called map field) is simplified as compared to fuzzy ARTMAP, and is therefore easier to implement without losing training flexibility. Also developed in this work is the capability of slow learning for fuzzy LAPART, with fast-commit and slow-recode options. For efficient coding of noisy input data, it is useful to use fast commitment and slow recoding training. With these upgraded features included in the LAPART architecture, the fuzzy LAPART neural network is capable of handling a wider class of problems not addressable with LAPART.

To illustrate the enhanced capability of fuzzy LAPART, a standard benchmark test was used which involved separation of spirals. The performance results were compared to two different counterpropagation networks. For every case run, the results showed that fuzzy LAPART outperformed counterpropagation. In the case of fuzzy LAPART, performance levels of 100%, 99.22% and 91.4% were achieved for baseline vigilance parameters of 1, 0.95 and 0.9, respectively. Fuzzy LAPART conjoins LAPART and fuzzy ARTMAP, however, for sequenced analog inputs to the neural networks, fuzzy LAPART is a preferred architecture. An extended investigation of an enhanced weight update methods between fuzzy LAPART's A and B fields could provide a basis for extensions to multi-module architectures for data fusion problems.

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*, vol. 37, pp. 54-115, 1987.
[2]   G. A. Carpenter and S. Grossberg, "ART2: Self-organization of Stable Category Recognition Codes for Analog Input Patterns," *Applied Optics*, vol. 26, pp. 4919-4930, 1987.

[3]   G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: An adaptive resonance algorithm for rapid, stable classification of analog patterns," *Proc. of IJCNN*, vol. II, pp. 411-420, 1991.

[4]   G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Networks*, Vol. 3, No. 5, pp. 698-713, 1992.

[5]   M. J. Healy, T. P. Caudell, and S. D. G. Smith, "A Neural Architecture for Pattern Sequence Verification Through Inferencing," *IEEE Trans. Neural Networks*, vol. 4. no.1, pp. 9-20, 1993.

[6]   L. A. Zadeh, "Fuzzy Logic," *IEEE Computer Magazine , pp. 83-93, 1988.*

[7]   B. Moore, "ART1 and pattern clustering," in *Proc. 1988 Connectionist Models Summer School,* 1989, pp. 174-185.

[8]   K. J. Lang and M. J. Witbrock, "Learning to tell two spirals apart," in *Proc. 1988 Connectionist Models Summer School,* pp. 52 - 59, 1989.

[9]   R. Hecht-Nielsen, "Counterpropagation Networks," *IEEE First International Joint Conference on Neural Networks, Washington. D.C.,* vol. I, pp. 593-605, 1987.

[10] L.V. Fausett, *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1993.