Euclidean ART Neural Networks

Riyadh Kenaya, *Member, IEEE* and Ka C Cheok, *Member, IEEE*

***Abstract*—Unsupervised Neural networks are known for their ability to cluster input vectors (patterns) into categories (neurons) based on a neighborhood of similarity between two or more patterns and how big the radius of similarity to bet set by the network user. Fuzzy ART neural networks are examples of such systems where normalized input patterns are clustered into categories. It has been proven however that fuzzy ART networks show a disorder in their clustering performance especially when they are trained to learn noisy patterns. While Fuzzy ART networks employ the fuzzy AND neighborhood to determine the pattern pertinence and category learning, Euclidean ART networks employ the Euclidean neighborhood to decide the said pertinence and patterns mean value for category training. Euclidean ART neural network or better known as EART is trained according to a certain algorithm that calculates the Euclidean distance and decides to whether include a new pattern in an already existing category (cluster) and update its position in the clustering map, or consider it as a new category if it is far enough from all of the existing categories. The abovementioned algorithm is tested in clustering patterns of certain distribution in the plane. Clustering results were collected for both fuzzy and Euclidean ART network for comparison purposes and both of the networks were retrained to learn noisy patterns in order to test their performance against noise.**

***Index Terms*— Fuzzy Adaptive Resonance Theory, Euclidean Adaptive Resonance Theory, Neural Networks, Pattern Clustering**

I. Introduction

Fuzzy ART [1], [2], [3] is a self organized clustering tool that is used to group huge sets of data into separate categories. ART uses a minimum required similarity between patterns that are grouped within one cluster. The resulting number of clusters then depends on the distance between all input patterns presented to the network during the training phase. This similarity is decided by a parameter to be set by the network trainer, and is known as vigilance parameter, ρ. ρ ranges between 0 and 1. Input patterns are usually created by converting the input signal into normalized vectors of M dimensions each. Once the input pattern is created, it should be compared with all of the N stored categories in the fuzzy ART memory. If the degree of similarity between the current input pattern and one of the best fitting categories is at least as high as vigilance ρ, the pattern would be added to the category cluster followed by modifying the category itself [4]. Modifying the said category is referred to as learning [5],[6],[7]. If the similarity value is less than the vigilance factor for all of the best fitting categories, a new cluster is to be added to the category library and the added cluster's first member is the pattern itself. Once one of the existing committed clusters matches the input well enough, we say that resonance [8] is reached and the category that best matches the input pattern is adapted through the above mentioned learning process. Learning process is represented by slightly shifting the category value towards the value of the said pattern vector. This process would continue as long as input patterns are presented to the fuzzy ART network and there is enough memory to be occupied by the added categories. It is very important to refer to the vigilance factor as a network critical factor, because it decides the number of categories and the radius of similarity of each category. A network with small vigilance factor witnesses a small number of categories with large neighborhood of similarity, whereas the network with a big vigilance value witnesses a bigger number of categories and small neighborhood of similarity [1]. As we increase the vigilance value, the network becomes more sensitive to small changes in input patterns and this would consequently result in increasing the number of resulting categories. The performance of the neural network is measured by its ability to cluster patterns that are similar to each other into one cluster and create a new cluster only when the pattern is far enough from clusters that have been created. It is an important issue that the neural network keeps the number of clusters as minimal as possible to prevent what is called proliferation of categories [2]. It is the case where the number of categories increases without bound. This would degrade the availability of the memory needed to store future categories. As a matter of fact, fuzzy ART networks show an acceptable and sound performance when the patterns presented to them are free of noise. This performance however is significantly affected when those patterns become noisy. The network would start to suffer the proliferation and category distribution disorder.

II. Fuzzy ART Structure and Algorithm

Fuzzy ART neural network can be characterized into three steps: Preprocessing, Searching (choice and match), and Adaptation levels [3]. These levels are described in below:

A. Preprocessing Step Based on Complement Coding Level

Carpenter and Grossberg have shown that input patterns should be normalized in a way that their vector components would be in the range [0,1]. Beside the said normalization process, those normalized patterns should be further complementally coded. The process of normalization and complement coding would ensure that the network would stay stable during the training process and that the cluster proliferation is prevented [1].

Each input pattern is an M-dimensional vector (I1, , , , , I_M), where each component Ii is in the interval [0,1]. [1].

$$\mathbf{I} = \frac{\mathbf{a}}{\|\mathbf{a}\|} \tag{1}$$

where

$$\| \mathbf{a} \| = \sum_{i=1}^{N} | a_i |, \text{ N is the length of vector } \mathbf{a} \quad (2)$$

Input patterns are further complementally coded. Complement coding represents both the on-response and off-response to a. [1]

So If **a** represents the on-response, the off-response is represented by $\mathbf{a}^c$ where

$$a_i^c = 1 - a_i \quad (3)$$

B. *Search* Level

The normalized pattern is to be compared with all of the saved categories found in the fuzzy ART memory. The comparison is based on calculating a similarity between the normalized input pattern and all of the saved categories. The function is known as activation function T.

Activation function Tj for each ART input pattern is calculated as in equation (4)

$$T_j = \frac{\| \mathbf{I} \wedge \mathbf{w}_j \|}{\alpha + \| \mathbf{w}_j \|} \quad (4)$$

Where **I** is the input pattern currently applied to fuzzy ART module and $\mathbf{w}_j$ is the jth category found in the ART module's memory, $\wedge$ is the fuzzy AND operation, (X $\wedge$ Y) = min(x_i , y_i) [9]. α is a constant > 0. J is the index that represents the category of which the activation function is highest. The category that wins the search is prepared to undergo a test called resonance or vigilance test.

C. *Vigilance Test Level*

Category $\mathbf{w}_J$ that resulted from the search level is now compared with the input pattern **I** (5).

$$\frac{\| \mathbf{I} \wedge \mathbf{w}_J \|}{\| \mathbf{I} \|} > \rho \quad (5)$$

Vigilance test is performed in order to measure the match of the category whose activation is highest with the input pattern **I**. If the match function exceeds the vigilance parameter ρ, we say that the Jth category has won and resonance is reached [2], otherwise the said category is excluded from the search for this pattern. The search process resumes looking for another category that maximizes the activation function and satisfies the vigilance test (5). If no category is found to satisfy both of the activation and vigilance conditions, a new category is formed and added to ART memory. The new formed category would be the input pattern **I** itself.

D. *Adaptation level*

In case of resonance, learning rule (6) is used to modify the category weights.

$$\mathbf{w}_J \text{ new} = \beta (\mathbf{I} \wedge \mathbf{w}_J \text{ old}) + (1 - \beta) \mathbf{w}_J \text{ old} \quad (6)$$

Where $\beta \in (0,1]$ and is known as learning rate [10]. When β =1, it is the case of fast learning where the category vector is directly equated to the input pattern. It is notable that not the input pattern but the attended portions of it are learned. This causes detection of relevant feature groupings of the categories and focusing attention on these portions while trying to match new input patterns [11]. Learning in fuzzy ART algorithm can be interpreted geometrically as the extension of the category region towards the input sample. The vigilance parameter controls the similarity to the input sample required from the chosen category, thus lowering the vigilance provides broader generalization. Fig. 1 shows the fuzzy ART algorithm flowchart.

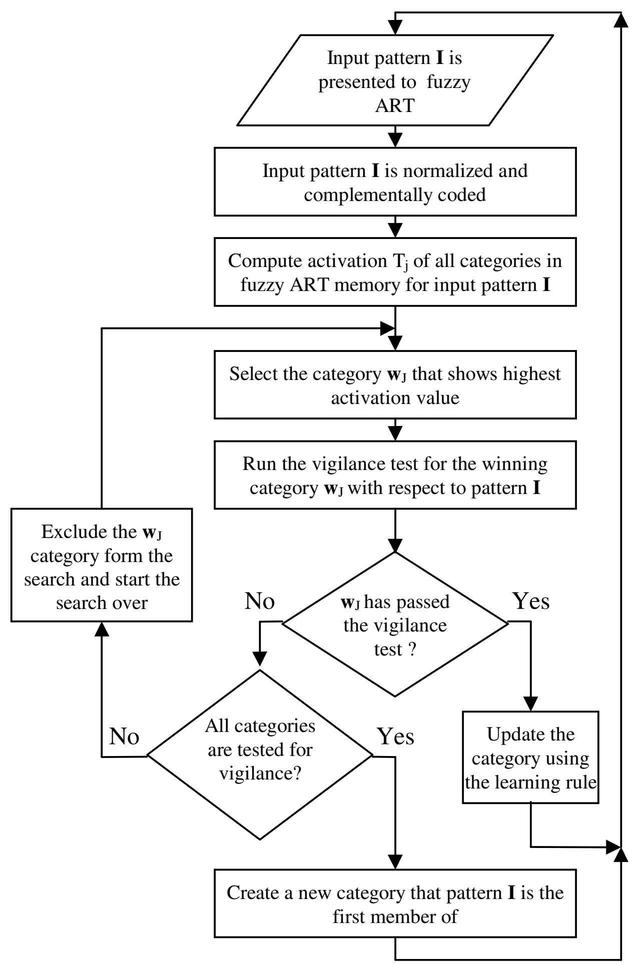

Fig.1 Fuzzy ART Algorithm Flowchart

III. EUCLIDEAN ART STRUCTURE AND ALGORITHM

Williamson has recognized two weaknesses of the Fuzzy ART: sensitivity to noise, and inefficiency representation of fuzzy categories [2]. To solve the problem we have to think of a learning scheme that tries to suppress the noise found in the pattern background. It is well known that the white Gaussian noise is an indeterministic signal of an average value of 0 [12]. This means that if we are able to make the fuzzy ART training process depends on averaging the patterns that are members of the same cluster, this would definitely result in suppressing the noise that shows up in the pattern background without the need to use a filter, i.e. the noise filter becomes inherent in the training process itself. This crucial matter has inspired us to look for a training scheme that replaces the fuzzy AND operation with an averaging operation.

A. *Euclidean ART Algorithm*

The Euclidean ART (EART) is a clustering technique that evaluates the Euclidean distance between patterns and cluster centers to decide pattern's clustering membership. This technique replaces the fuzzy operations found in fuzzy ART with the Euclidean distance evaluation. The learning rule is an averaging process used to calculate the new cluster center location after a new pattern is added to the cluster.

The pattern membership is dependent on the parameter R_{th}, the Euclidean threshold. This parameter determines the cluster radius of neighborhood which decides the pattern membership. If the presented pattern lies within this neighborhood, the pattern is considered as a cluster member. R_{th} factor is similar to the vigilance factor ρ found in fuzzy ART networks. The smaller the R_{th} value, the larger the number of clusters –categories– to result form the training process and the smaller the number of pattern memberships per category. This is equivalent to a fuzzy ART network with high vigilance factor. We can summarize the EART algorithm in the following steps:

Step 1: Present a normalized and complementally coded pattern to EART module. The normalization algorithm is explained later.

Step 2: Calculate the Euclidean distance between this pattern and the entire existing cluster centers (7). Those Euclidean distances are considered as an activation value of each cluster center with respect to the presented pattern. If there is no cluster center yet, consider this pattern to be the first one.

$$d(j) = \sum_{j=1}^{N} \sqrt{\left(\mathbf{x}_i - \mathbf{w}_j\right)^2} \tag{7}$$

where j is the category index found in EART network and i is the index of the current presented pattern.

Step 3: Find d(J), where d(J) = min(**d**).

Step 4: **If** d(J) $\leq R_{th}$ **then**

-Include the presented pattern $\mathbf{x}_k$ in the winning cluster whose center is $\mathbf{w}_J$.

-Start the learning process; calculate the new cluster center according to learning equation (8).

$$\mathbf{w}_J \text{ new} = \frac{\sum_{k=1}^{L} \mathbf{x}_{Jk}}{L} \tag{8}$$

where $\mathbf{x}_{Jk}$ is the pattern member k of cluster J, L is the number of cluster members

Else $\mathbf{x}_i$ becomes a new category $\mathbf{w}_{N+1}$.

Step 5: Jump back to step 1 to accept a new pattern if there are more patterns to test. Else the training is over and the resulting EART matrix is the trained EART network.

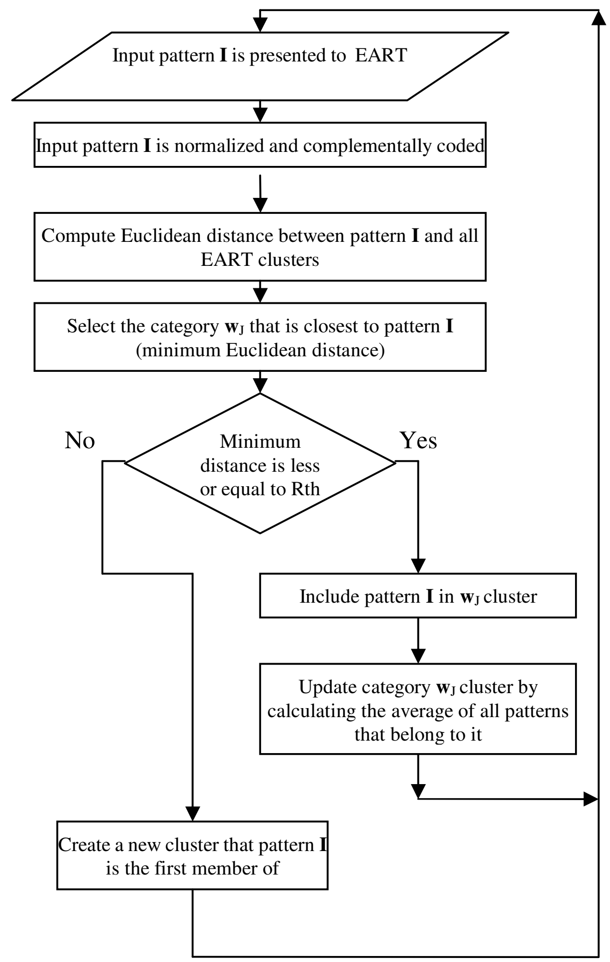

Fig.2 Euclidean ART Algorithm Flowchart

B. *Euclidean ART Noise Analysis*

As mentioned earlier, fuzzy ART networks show poor performance in learning noisy patterns. This means that the locations of cluster centers would be highly affected by the noise background that accompanies the presented pattern. Noise problem becomes a minor problem when it comes to train EART network. Since training is dependent on averaging the cluster members, the noise added to those members would undergo same averaging process. We can better describe the process mathematically in equation (9).

Let the noisy pattern presented to the EART network be represent as

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i + \mathbf{n}_i \quad (9)$$

where $\tilde{\mathbf{x}}_i$ is the noisy pattern vector presented to the EART network and consists of the original pattern vector $\mathbf{x}_i$ added to it the background white noise vector $\mathbf{n}_i$. When the conditions of learning are fulfilled, training equation (10) calculates the new cluster center location, so

$$\mathbf{w}_J \text{ new} = \frac{\sum_{k=1}^{L} \tilde{\mathbf{x}}_{Jk}}{L} \quad (10)$$

then

$$\mathbf{w}_J \text{ new} = \frac{\sum_{k=1}^{L} \mathbf{x}_{Jk}}{L} + \frac{\sum_{k=1}^{L} \mathbf{n}_{Jk}}{L} \quad (11)$$

It is obvious that equation (11) right hand side terms are the expected values of the cluster members and the noise vectors respectively. It is also known that the expected value of any random signal over a huge collection of points is 0, which in our case applies to noise vectors.
So we can easily conclude that

$$\mathbf{w}_J \text{ new} = \frac{\sum_{k=1}^{L} \mathbf{x}_{Jk}}{L} \quad (12)$$

for large number of training patterns L. We see that EART training equation plays an important role noise suppression.

IV. Euclidean ART Training and Calcification Results

We can test the EART algorithm and compare its performance with fuzzy ART algorithm performance for a number of pattern distributions. Our test and comparison strategy is represented by clustering a collection of data patterns in the plane and locate their cluster centers. Patterns used are tow dimensional vectors in the plane. Noisy patterns are then used to train both of the fuzzy ART and EART systems.

A. Input pattern Normalization and Complement Coding

Normalization can be summarized in the flowing steps

1- Find maximum and minimum values of x and y components in their corresponding pattern vectors.
2- Evaluate the normalization ranges by finding the difference between maximum and minimum values for x and y respectively.

$$\text{axis range} = \text{axis max value} - \text{axis min value} \quad (13)$$

3- Shift axis values up by adding the absolute value of the minimum axis value if it is negative. Otherwise don't shift.
4- Calculate axis normalized values by dividing the shifted axis values by axis range.

Fig. 3 and 4 shows spiral distribution for normalized and un-normalized patterns in the plane. The normalization algorithm modifies the collected patterns base scale they belong to and converts it to a normalized scale that ranges between zero and one.

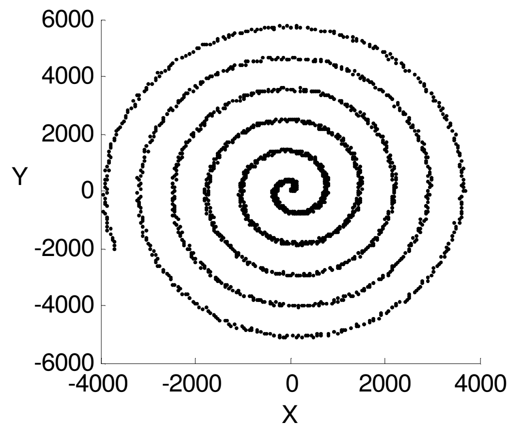

Fig.3 Un-normalized pattern distribution in the X-Y plane.

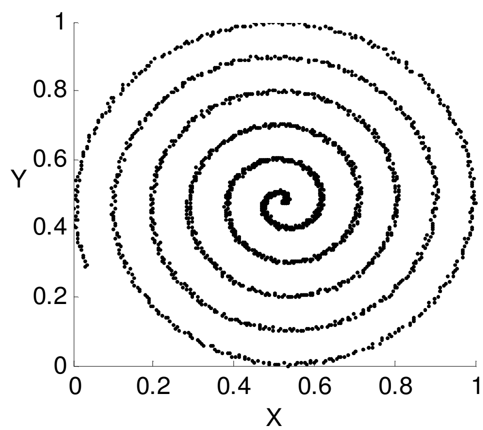

Fig.4 Normalized pattern distribution in the X-Y plane.

B. Training Results for Clean Patterns

We suggest different pattern distributions in the plane and train both Euclidean and fuzzy ARTMAP networks for those distributions. This is a convenient way to see how each system responds. We can also compare the number of resulting categories and their distribution.

Fig. 5, 6, 7, and 8 shows cluster and pattern maps for both of fuzzy and Euclidean ART networks. Dots represent patterns and the crosses represent the cluster centers (categories). By inspecting the cluster distribution, we can see how homogeneous the Euclidean ART clustering is in comparison with that of the fuzzy ART. By homogenous distribution we mean that patterns within a certain neighborhood belong to the same cluster and no pattern is left behind without a cluster membership, or at least no pattern is included in a cluster that represents features which are not close to the features of the pattern itself. On the other hand non-homogeneous clustering is the clustering where one or more patterns may belong to clusters of different features or left unclustered. The state of clustering and its being homogeneous or non homogeneous is highly dependent on the neighborhood factor of the cluster itself, which are the vigilance parameter ρ in fuzzy ART and the threshold radius Rth in EART networks.

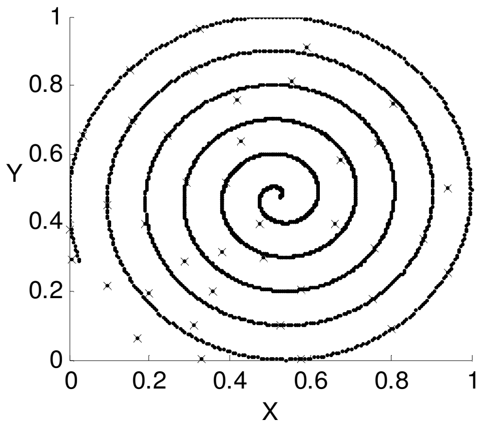

Fig.5 Fuzzy ART Clustering, ρ = 0.85, L = 2000, N = 39

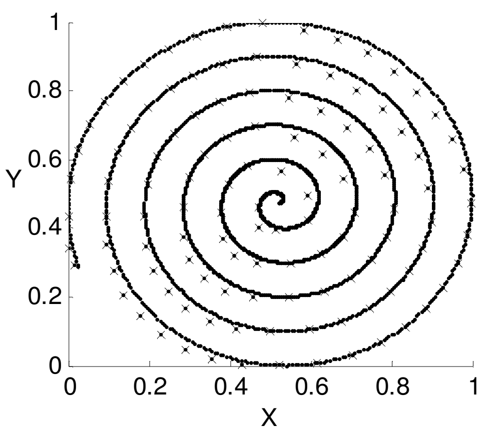

Fig.6 Fuzzy ART Clustering, ρ = 0.95, L = 2000, N = 116

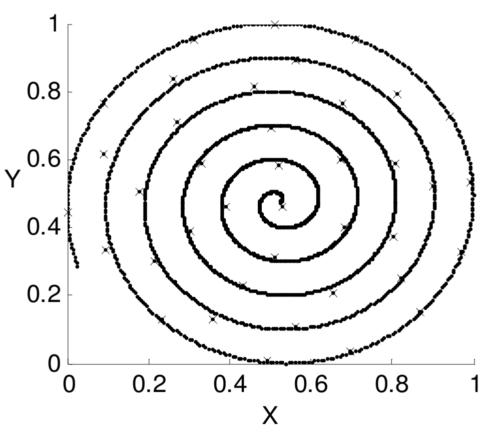

Fig.7 Euclidean ART Clustering Rth = 0.1518, L = 2000, N = 39

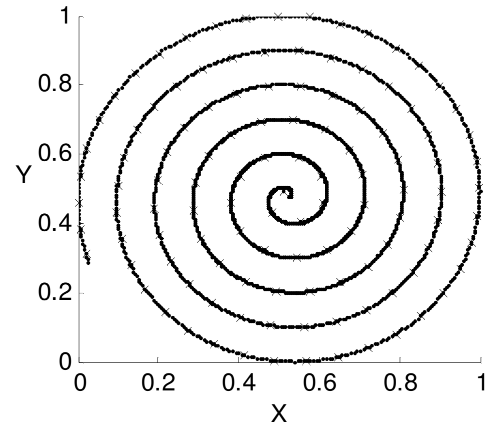

Fig.8 Euclidean ART Clustering Rth = 0.0605, L = 2000, N = 116

C. Training Results for Noisy Patterns

We can study the effect of noise on both of fuzzy and Euclidean ART networks, by adding a random signal vector -noise vector- to the training patterns and monitor the clustering behavior of both of the networks. Fig. 9, 10, 11, and 12 shows how EART network is immune against the added noise.

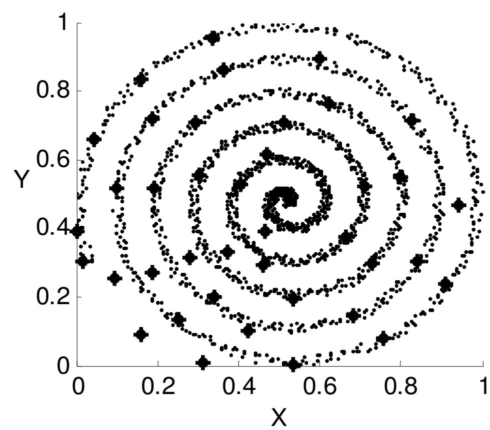

Fig.9 Fuzzy ART Clustering ρ = 0.85, L = 2000, N = 40

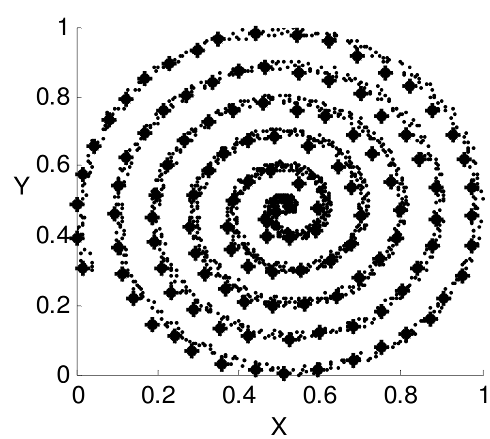

Fig.10 Fuzzy ART Clustering ρ = 0.94963, L = 2000, N = 119

ISBN: 978-988-98671-0-2

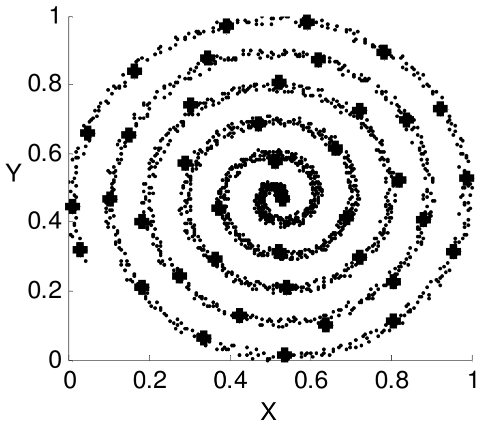

Fig.11 Euclidean ART Clustering, Rth = 0.159, L = 2000, N = 40

By comparing figures 9 and 11, we can easily tell that EART is doing a better job in clustering noisy patterns into categories that are still located at the center of the pattern

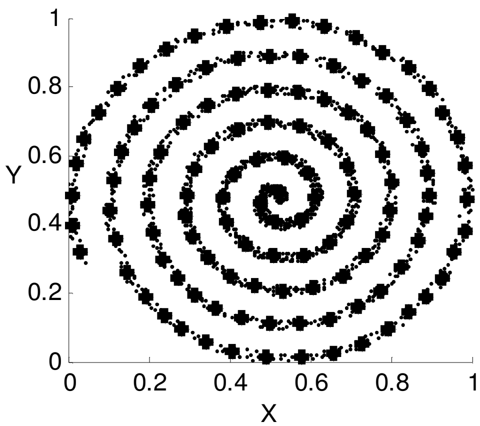

Fig.12 Euclidean ART Clustering, Rth = 0.0652, L = 2000, N = 119

collection, whereas fuzzy ART clustering is less homogeneous and we can see some bare areas where the patterns are not clustered into a category. Homogeneous distribution of categories is very important to ensure the best classification performance of the network. Networks whose patterns are left uncategorized during the training phase or at least included in a cluster that belong to a far category would fail the classification phase. Such failure would result in assigning the pattern to be classified to the wrong category. Fuzzy ART vigilance parameter ρ plays an important role in deciding the number of categories and how are they distributed. Fig. 5 shows fuzzy ART category distribution for low vigilance value of 0.85 which resulted in 39 poorly distributed categories. The outer circle of the spiral witnesses a bare area and the inner circles witness categories deviated from the pattern collection. Increasing the vigilance parameter improves the category distribution (fig 6) by covering the bare areas but at the same time increases the number of the resulting categories and does not completely clear the category deviation problem. Having the category distribution associated with the value of vigilance parameter, weakens the performance of the network because of the proliferation problem that accompanies the increase of vigilance value [13]. Those problems however are completely solved by EART architecture where the category distribution is independent of the neighborhood parameter Rth as shown in fig. 7 and 8. EART clusters shown in fig. 7, 8, 11, and 12 are homogeneously distributed and occupy the center of their belonging patterns. The said figures show that the category distribution and location in EART is less dependent on the neighborhood parameter Rth.

V. CONCLUSION

It was also very clear that the noise effect is highly suppressed when Euclidean ART network is used for clustering noisy patterns by virtue of the averaging process that is used to calculate the new cluster position during training. This is how Euclidean ART neural network surpasses its fuzzy ART counterpart. Our future work will be the supervised version of the Euclidean ART algorithm, the Euclidean ARTMAP. The importance of the proposed system comes from the necessity to overcome the obstacles and limitations that fuzzy ARTMAP neural networks faces. Our future work will also consider training Euclidean ARTMAP systems to imitate the human behavior. This could be done via designing robotic systems and train their Euclidean ART controllers to respond in the same way the man responds to an external stimulation.

REFERENCES

[1] G. A. Carpenter, S. Grossberg, and D. B. Rosen. "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system." *IEEE Transactions on Neural Networks, V4: 759-771, 1991.*

[2] H. Xu. *Mahalanobis Distance-Based ARTMAP Networks*, Master's thesis of Computer Science, San Diego State Univ., 2003.

[3] A. Akhbardeh, and A. Värri. "Novel Supervised Fuzzy Adaptive Resonance Theory (SF-ART) Neural Network for Pattern Recognition". *Intelligent Signal Processing, 2005 IEEE International Workshop on Volume, Issue, 1-3 Sept. 2005 Page(s): 149 – 154.*

[4] T. Frank, K. F Kraiss, and T. Kuhlen. "Comparative Analysis of Fuzzy ART and ART-2A Network Clustering Performance". *IEEE Transaction on Neural Network, V9:* 544-559, 1998.

[5] H. Ueda, H. Kimoto, T. Naraki, K. Takahashi, and T. Miyahara. "Categorization of Continuous Numeric Percepts By Modified Fuzzy Art With Q Learning". *International Journal of Pattern* Recognition *and Artificial Intelligence, Vol. 20, No. 2 (2006) 113–127.*

[6] A. Pekery and Y. Karay. "Parameter Setting of The Fuzzy ART Neural Network to Part–Machine Cell Formation Problem*". International Journal of Production Research 2004, vol. 42, no. 6, 1257–1278.*

[7] N. C. Suresht and S. Kaparthit. "Performance of Fuzzy ART neural network for group technology cell formation". *International Journal of Production Research, 1994, VOL. 32, NO. 7, 1693-1713.*

[8] A. P. Uribe. *Structure-Adaptable Digital Neural Networks,* PhD thesis, Federal polytechnic school of Luazzane, Switzerland, 1999.

[9] J.S.R. Jang, C.T. Sun, E. Mizutani. *Neuro-Fuzzy and Soft Computing*, Prentice Hall, Upper Saddle River, NJ 1997.

[10] G. A. Carpentert, Stephen Grossberg, and John H. Reynolds. "ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network." *Nurall Nelworks, Vol, 4. pp, 565-588. 1991.*

[11] O. C. Ozcanli – Ozbay, *An Image Retrieval System Based on Region Classification,* Master's thesis of Computer Science, Middle East Technical University, Turkey, 2004.

[12] M. Shinsaku, S Iwao, and T. Kiyoshi, "Duplex Adaptive Digital Filter and Method of Operation". *United States Patent 5535149, Publication Date:07/09/1996* .

[13] H. Isawa, M. Tomita, H. Matsushita and Yoshifumi Nishio, "Fuzzy Adaptive Resonance Theory with Group Learning and its Applications", 2007 *International Symposium on Nonlinear Theory and its Applications NOLTA'07, Vancouver, Canada, pp,292-295. September 16-19, 2007.*