Fuzzy Adaptive Resonance Theory Combining Overlapped Category in Consideration of Connections

Haruka Isawa, Haruna Matsushita and Yoshifumi Nishio

***Abstract*— Adaptive Resonance Theory (ART) is an unsupervised neural network. Fuzzy ART (FART) is a variation of ART, allows both binary and continuous input patterns. However, Fuzzy ART has the category proliferation problem. In this study, to solve this problem, we propose a new Fuzzy ART algorithm: Fuzzy ART Combining Overlapped Category in consideration of connections (C-FART). C-FART has two important features. One is to make connections between similar categories. The other is to combine overlapping categories into with connections one category. We investigate the behavior of C-FART, and compare C-FART with the conventional FART.**

I. Introduction

SELF-ORGANIZED clustering is a powerful tool whenever huge sets of data have to be divided into separate categories. In the field of neural network, the Adaptive Resonance Theory (ART), introduced and developed by G.A. Carpenter and S. Grossberg [1], is a popular representative for self-organized clustering. Some outstanding features of ART, besides its clustering capabilities, have attracted the attention from application engineers. This theory has evolved as a series of real-time neural network models that perform unsupervised and supervised learning, pattern recognition, and prediction. These models are capable of learning stable recognition categories in response to arbitrary input sequences. However, the fusion of a computational efficiency of neural network and a capability of fuzzy logic to represent complex class boundaries, has created a lot of interest in neurofuzzy pattern recognition systems. Then, we pay attentions Fuzzy ART (FART) [2], which is the merger of fuzzy logic and ART neural network. FART is an unsupervised neural network and incorporates the basic features of all ART systems [3]-[4]. Furthermore, by creating hyperboxes, input vectors are classified into each appropriate category. For this reason, FART often makes input data of the common categories classify several categories, and FART performance is highly dependant on a vigilance parameter, which controls hyperbox size. Therefore, FART causes the category proliferation problem. In order to solve this problem, the method considering overlapping categories is proposed in hyperbox fuzzy set [6]. However, if all overlapping categories are combined in FART, the input data are not classified into each appropriate category since the hyperboxes keep getting larger and larger.

Our past study proposed Fuzzy ART with Group Learning (FART-GL) [5]. FART-GL has connections between categories as human relationship which keeps changing with time in the real world. The connection is created between similar categories, and the categories, which have connections, are learned as "group" of category. By using this method, sample data are effectively classified into each appropriate group. However, FART-GL can not reduce category proliferation.

Haruka Isawa, Haruna Matsushita and Yoshifumi Nishio are with the Department of Electrical and Electronic Engineering, Tokushima University, Tokushima, 770–8506, Japan (phone: +81–88–656–7470; fax: +81–88–656–7471; email: {haruka, haruna, nishio}@ee.tokushima-u.ac.jp).

In this study, we propose a Fuzzy ART Combining Overlapped Category in consideration of connections (C-FART) to solve the category proliferation problem. C-FART makes connections between similar categories or releases connections at each step, in addition, C-FART combines the overlapping categories using the created connections. In other words, C-FART combines the categories with due consideration of their similarity.

In Section II, the algorithm of the conventional FART is introduced. In Section III, we explain the learning algorithm of the proposed C-FART in detail. We apply C-FART to various input data sets, and the learning behaviors of C-FART are investigated in Section IV. The learning performance of C-FART is compared with FART. We confirm that the proposed C-FART can perform more effective learning than the conventional FART and can reduce the category proliferation problem.

II. Conventional Fuzzy ART

A. Structure of Fuzzy ART

Fuzzy ART is composed of F_1 (input layer) and F_2 (category layer). F_1 and F_2 are connected by the bottom-up-weight vector $\boldsymbol{w}_{ij}$ and the top-down-weight vector $\boldsymbol{w}_{ji}$. m neurons of the input layer F_1 correspond to the an input vector $\boldsymbol{I}$.

Input vector: Each input $\boldsymbol{I}$ is an m-dimensional vector $\boldsymbol{I} = (I_1, I_2, \cdots, I_m)$, where each component I_i $(i = 1, \cdots, m)$ is in the interval [0, 1].

Weight vector: Each category j corresponds to a vector $\boldsymbol{w}_j = (w_{j1}, \cdots, w_{jm})$, $(j = 1, \cdots, n)$ of adaptive weight, or LTM (long-term-memory) traces. The number of potential categories n is arbitrary. Initially

$$w_{j1} = \cdots = w_{jm} = 1. \tag{1}$$

Parameters: Fuzzy ART dynamics are determined by *choice parameter* $\alpha > 0$; *learning parameter* $\beta \in [0, 1]$; and *vigilance parameter* $\rho \in [0, 1]$.

B. Learning Algorithm of Fuzzy ART

We explain the learning algorithm of the conventional Fuzzy ART.

978-1-4244-1821-3/08/$25.00©2008 IEEE

(FART1) An input vector $\boldsymbol{I}$ is inputted to the category layer F_2 from the input layer F_1.
(FART2) A winning category J is chosen. For the input vector $\boldsymbol{I}$ and category j, *choice function* T_j is defined by

$$T_j(\boldsymbol{I}) = \frac{|\, \boldsymbol{I} \wedge \boldsymbol{w}_j \,|}{(\alpha + |\, \boldsymbol{w}_j \,|)}, \tag{2}$$

where the fuzzy AND [7] operator $\wedge$ and the norm $|\cdot|$ are defined by

$$(\boldsymbol{p} \wedge \boldsymbol{q})_i \equiv \min(p_i, q_i), \tag{3}$$

$$|\, \boldsymbol{P} \,| \equiv \sum_{i=1}^{m} |\, p_i \,|. \tag{4}$$

The winning category J, whose T_j is maximum, is found;

$$T_J = \max\{T_j : j = 1 \cdots n\}. \tag{5}$$

If more than one T_j is maximal, the category j with the smallest index is chosen as the winner J.
(FART3) The similarity of $\boldsymbol{I}$ and the current winning category $\boldsymbol{w}_J$ is measured by the vigilance criterion. We check whether

$$\frac{|\, \boldsymbol{I} \wedge \boldsymbol{w}_J \,|}{|\, \boldsymbol{I} \,|} \geq \rho. \tag{6}$$

If Eq. (6) is not satisfied, namely

$$\frac{|\, \boldsymbol{I} \wedge \boldsymbol{w}_J \,|}{|\, \boldsymbol{I} \,|} < \rho, \tag{7}$$

a new index J is chosen by Eq. (5). The search process continues until the chosen J satisfies Eq. (6).
(FART4) $\boldsymbol{w}_J$ is updated by

$$\boldsymbol{w}_J^{\text{new}} = \beta(\boldsymbol{I} \wedge \boldsymbol{w}_J^{\text{old}}) + (1-\beta)\boldsymbol{w}_J^{\text{old}}, \tag{8}$$

if there is J which satisfies Eq. (6). On the contrary, if all available F_2 nodes do not satisfy Eq. (6), a new category is established in F_2;

$$\boldsymbol{w}_{n+1} = \boldsymbol{I}. \tag{9}$$

(FART5) The steps from (FART1) to (FART4) are repeated for all the input data.

C. Complement Coding

Because vector elements of prototype can only become smaller by adaptation, a Fuzzy ART network tends to create more and more prototype over time. This behavior is avoided by normalizing input to constant vector length, this method is called *complement coding*. The complement coded input $\boldsymbol{I}$ to the recognition system is the $2m$-dimensional vector;

$$\boldsymbol{I} = (\boldsymbol{a}, \boldsymbol{a}^c) \equiv (a_1, \cdots, a_m, a_1^c, \cdots a_m^c). \tag{10}$$

where

$$\boldsymbol{a}_i^c \equiv 1 - a_i. \tag{11}$$

In this case, the weight vector $\boldsymbol{w}_j$ can be written in complement coding form;

$$\boldsymbol{w}_j = (\boldsymbol{u}_j, \boldsymbol{v}_j^c) = (w_{j1}, \cdots, w_{jm}, w_{jm+1}, \cdots, w_{j2m}), \tag{12}$$

where $\boldsymbol{u}_j$ and $\boldsymbol{v}_j$ are m-dimensional vectors as $\boldsymbol{u}_j = (w_{j1}, \cdots, w_{jm})$ and $\boldsymbol{v}_j = (1 - w_{jm+1}, \cdots, 1 - w_{j2m})$.

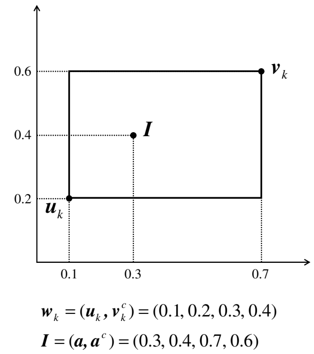


Fig. 2. If $\boldsymbol{I}$ is fully contained by any category k, $\boldsymbol{I}$ is classified to the category k even if k is not the winning category.

III. Proposed Fuzzy ART (C-FART)

In this study, we propose the combining overlapped category spaces Fuzzy ART considering connections (C-FART). C-FART learning algorithm consists of mainly three steps: 1) Learning, 2) Update Connections and 3) Combining Categories. A flowchart of the learning algorithm is shown in Fig. 1. As important features of C-FART, C-FART learns connections between categories, and overlapped categories are combined with considering their connections. Therefore, if there is no connection between categories, overlapping categories are not combined.
Connection: C-FART has a connection matrix denoted by C and the age of the connections denoted by age. Both C and age are $n \times n$ matrices where$n$ is the number of potential categories. The initial values of C and age are set to zero:

$$C = 0, \quad age = 0. \tag{13}$$

If the categories J and j are connected, $C_{J,j}$ changes from zero to one.

A. Learning Algorithm of C-FART

We explain the learning algorithm of C-FART in detail.

Learning Steps
(C-FART1) An input vector $\boldsymbol{I}$ is inputted to the category layer F_2 from the input layer F_1.
(C-FART2) A winning category J is chosen according to the step (FART2). Furthermore, a second-winning category J_2, whose T_{J_2} is the second largest next to T_J, is found for updating connections if J_2 exists.
(C-FART3) As the step (FART3), the similarity of the input $\boldsymbol{I}$ and the current winning category $\boldsymbol{w}_J$ is measured by the vigilance criterion of Eqs. (6) and (7).
(C-FART4) If any J satisfies Eq. (6), $\boldsymbol{w}_J$ is updated by Eq. (8) and we perform (C-FART5). If all the categories do not satisfy Eq. (6), we check whether $\boldsymbol{I}$ is fully contained by a category k (as Fig. 2). In other words, if $\{\min(w_{ki}, 1 - w_{km+i}) < a_i < \max(w_{ki}, 1 - w_{km+i})\}$ is

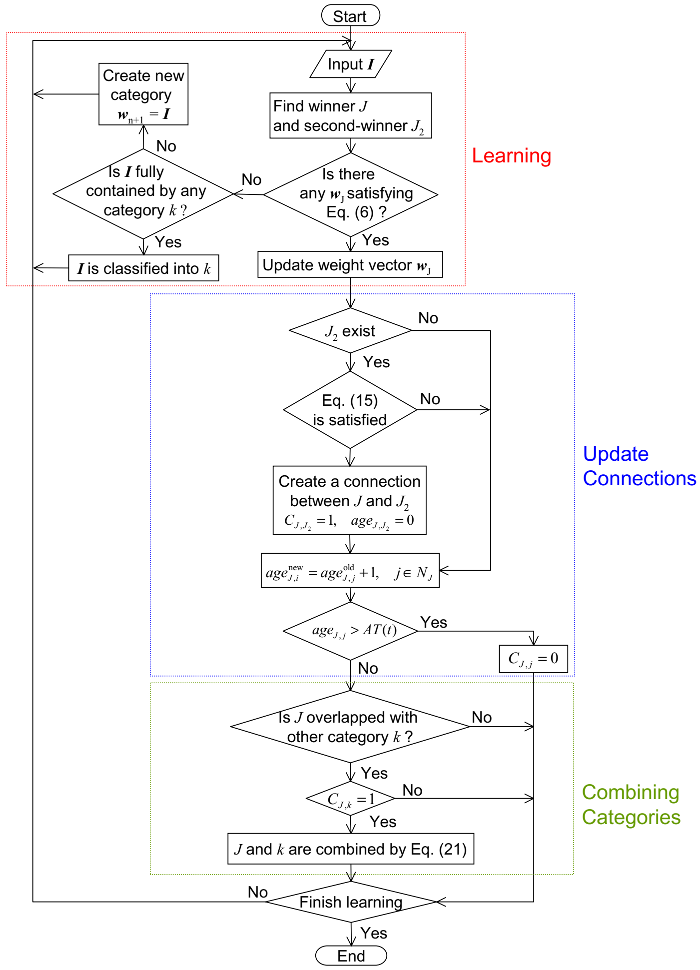


Fig. 1. Flowchart of learning algorithm of C-FART.

true for all dimension i, $\boldsymbol{I}$ is classified to the category k, $\boldsymbol{w}_k$ is updated;

$$\boldsymbol{w}_k^{\text{new}} = \beta(\boldsymbol{I} \wedge \boldsymbol{w}_k^{\text{old}}) + (1-\beta)\boldsymbol{w}_k^{\text{old}} \equiv \boldsymbol{w}_k^{\text{old}}, \tag{14}$$

and we perform (C-FART9) without the steps (C-FART5)-(C-FART8). If more than one categories satisfy Eq. (6), the category k with the smallest index is chosen. If $\boldsymbol{I}$ is fully contained by no categories, a new category is established according to Eq. (9) and we perform (C-FART9) without the steps (C-FART5)-(C-FART8).

Update Connections

(C-FART5) If J_2 does not exist, we skip this step and perform (C-FART6). The similarity of the input $\boldsymbol{I}$ and the second-winning category $\boldsymbol{w}_{J_2}$ is measured by

$$\frac{|\ \boldsymbol{I} \wedge \boldsymbol{w}_{J_2}\ |}{|\ \boldsymbol{I}\ |} \geq \rho. \tag{15}$$

If Eq. (15) is satisfied, a connection between the winning category J and the second-winning category J_2 is created;

$$C_{J,J_2} = 1. \tag{16}$$

The age of the connection between J and J_2 is set to zero ("refresh" the age);

$$age_{J,J_2} = 0. \tag{17}$$

On the contrary, if Eq. (15) is not satisfied, the connection is not updated

(C-FART6) The age of all categories, which directly connect with the winning category J, are increased one;

$$age_{J,j}^{\text{new}} = age_{J,j}^{\text{old}} + 1, \qquad j \in N_J, \tag{18}$$

where N_J is the set of categories which directly connect with J, namely $C_{J,j} = 1$.

(C-FART7) The connections are removed, if their age exceeds a threshold value $AT(t)$;

$$C_{J,j} = 0, \qquad age_{J,j} \geq AT(t), \tag{19}$$

where

$$AT(t) = AT_i \left(\frac{AT_f}{AT_i} \right)^{\frac{t}{t_{\max}}}, \tag{20}$$

where t is the learning step, $t_{\max}$ is the learning length, AT_i and AT_f is the initial value and the final value of AT, respectively.

Combining Categories

(C-FART8) We check whether the winning category J is overlapped with other category and combine these categories. We find the category k which are overlapping with J as Figs. 3(a) and (c), in other words, $\{\min(w_{Ji}, 1 - w_{Jm+i}) < w_{ki} < \max(w_{Ji}, 1 - w_{Jm+i})\}$ or $\{\min(w_{Ji}, 1 - w_{Jm+i}) < 1 - w_{km+i} < \max(w_{Ji}, 1 - w_{Jm+i})\}$ is true for any dimension i. If k exists and directly connects with J, namely $C_{J,k} = 1$, J and k are combined as Figs. 3(b) and (d) by

$$\boldsymbol{w}_J^{\text{new}} = \{(\boldsymbol{u}_J \wedge \boldsymbol{v}_J \wedge \boldsymbol{u}_k \wedge \boldsymbol{u}_k), (\boldsymbol{u}_J^c \wedge \boldsymbol{v}_J^c \wedge \boldsymbol{u}_k^c \wedge \boldsymbol{u}_k^c)\}. \tag{21}$$

The inputs belonging to the category k are classified to the category J, and the category k is removed, therefore, $n^{\text{new}} = n^{\text{old}} - 1$.

(C-FART9) The steps from (C-FART1) to (C-FART8) are repeated for all input data set. Therefore, C-FART makes connections or releases connections at each step, and the overlapped categories are combined with considering connections.

IV. Simulation Results

We apply the proposed C-FART to various input data and compare C-FART with the conventional FART.

A. Simulation 1

First, in order to confirm the behavior of the proposed C-FART, we consider simple 2-dimensional input data as Fig. 4(a), consisting of 2-clusters, whose distribution is non-uniform. Total number of the input data is 500 points, and the input data are sorted at random. 250 points are distributed within a rectangular range from 0.0 to 0.2 horizontally and from 0.0 to 0.6 vertically. The remaining 250 points are distributed within a rectangular range from 0.4 to 0.6 horizontally and from 0.0 to 0.6 vertically. Both FART and C-FART start learning with no categories, namely $n = 0$. The parameters for the learning are chosen as follows;

(For FART)

$$\alpha = 0.1, \ \beta = 1.0, \ \rho = 0.8,$$

(For C-FART)

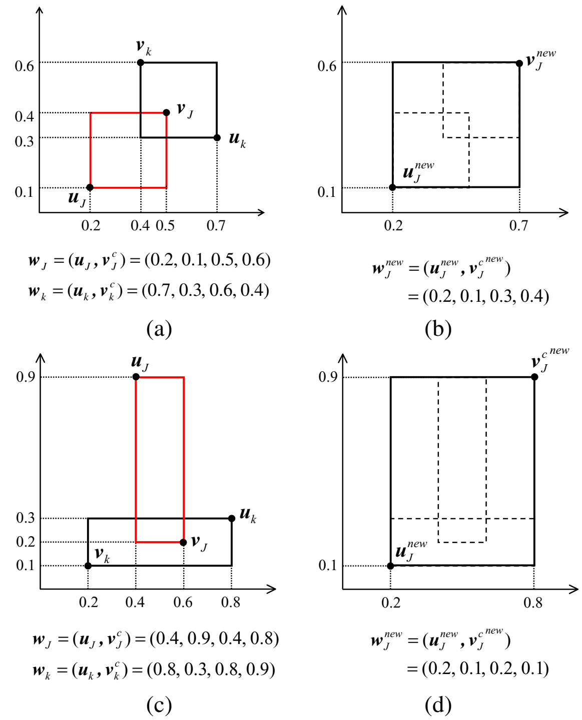


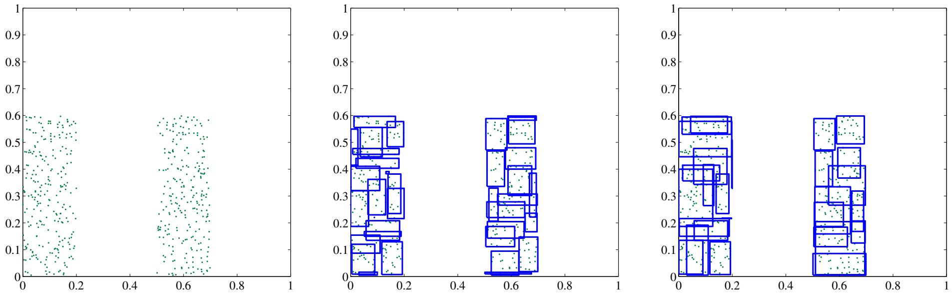
Fig. 3. Overlapped categories combining process in C-FART. (a) Category J and category k are overlapping. (b) If $C_{J,k} = 1$, $\boldsymbol{w}_J$ is updated according to Eq. 21, and k is removed. (c) Category J partially contains category k. (d) If $C_{J,k} = 1$, $\boldsymbol{w}_J$ and k are combined according to Eq. 21.

$$\alpha = 0.1, \ \beta = 1.0, \ \rho = 0.8, \ AT_i = 2, AT_f = 24,$$

where we use the same α, β and ρ for the conventional FART and the proposed C-FART for a fair comparison and the confirmation of the combining effect.

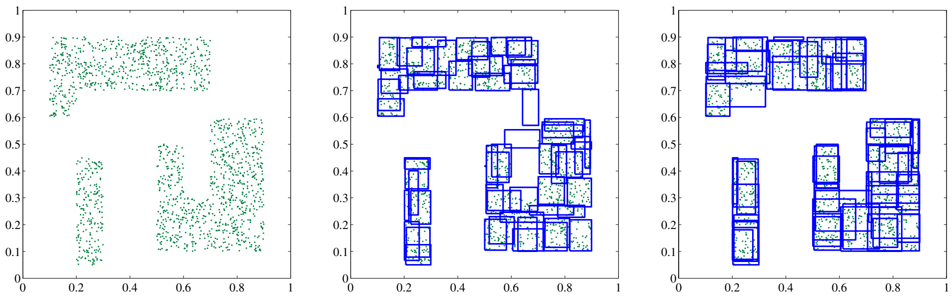
Learning results of the conventional FART and the proposed C-FART are shown in Figs. 4(b) and (c), respectively. We can see that the categories of C-FART are larger than FART. This is because the proposed C-FART combines the overlapping categories in consideration of their connections. If all overlapping categories are combined, the input data are not classified into each appropriate category since the hyperboxes keep getting larger and larger. However, the proposed C-FART makes connections between similar categories or releases connections at each step, in addition, C-FART combines the overlapping categories using the created connections. In other words, C-FART combines the categories with due consideration of their similarity. In the result of Fig. 4, the numbers of the categories of FART and C-FART are 34 and 28, respectively. The proposed C-FART reduces the category proliferation problem 20.00% from the conventional FART.

Furthermore, in order to appreciate the training performance of the two algorithms, we perform the learning simulation 100 times to the input data created in the same way as Fig. 4(a). The input data are created each time and



(a) (b) (c)

Fig. 4. Simulation of FART and C-FART for 2-clusters input data. (a) Input data. (b) Simulation result of conventional FART. (c) Simulation result of proposed C-FART.



(a) (b) (c)

Fig. 5. Simulation of FART and C-FART for 3-clusters input data. (a) Input data. (b) Simulation result of conventional FART. (c) Simulation result of proposed C-FART.

TABLE I
NUMBER OF CATEGORIES IN SIMULATION 1.

Algorithm	Number of categories		
	Minimum	Maximum	Average
FART	32	43	36.88
C-FART	24	37	31.17
Improvement rate	13.95%	25.00%	15.48%

are sorted by random each time to remove the dependence on the order of the input data.

Table I summarizes minimum, maximum and average values of the number of categories. In other words, the minimum and the maximum values mean best and worst results in 100 simulations, respectively. All the number of categories of the proposed C-FART are smaller than the conventional FART. From the average values, we can see that C-FART can reduce the category proliferation problem 15.48% from using the conventional FART. We can say that C-FART can obtain more effective result than FART.

B. Simulation 2

Next, we apply C-FART to 1650 points input data which have 3-clusters as Fig. 5(a). The top-left cluster has 650 points, the bottom-left cluster has 200 points, and the bottom-right cluster has 800 points. It is difficult to classify the input data such as Fig. 5(a) into appropriate categories because this data contains clusters whose size are different. The learning conditions are the same used in Simulation 1.

Figs. 5(b) and (c) show the learning results of the conventional FART and the proposed C-FART, respectively. We can see that the conventional FART classifies data points belonging to different clusters into same category. On the other side, the proposed C-FART can classify each input data into each appropriate category although the hyperbox size of C-FART is larger than FART.

Furthermore, we repeat this learning 100 times in the same way as Simulation 1 and examine the minimum, maximum and average values of the number of categories. The results are summarized in Table II. We can see that all the number of categories of the proposed C-FART are smaller than the conventional FART, and C-FART reduces categories by an average of 12.98%. This is because C-FART combines the overlapping categories with due consideration of their connections.

C. Simulation 3

Furthermore, we apply the proposed C-FART to rounded input data as Fig. 6(a). The input data has two clusters: a round shape cluster and a ring-shaped cluster whose distribution is non-uniform. The round shape cluster and the

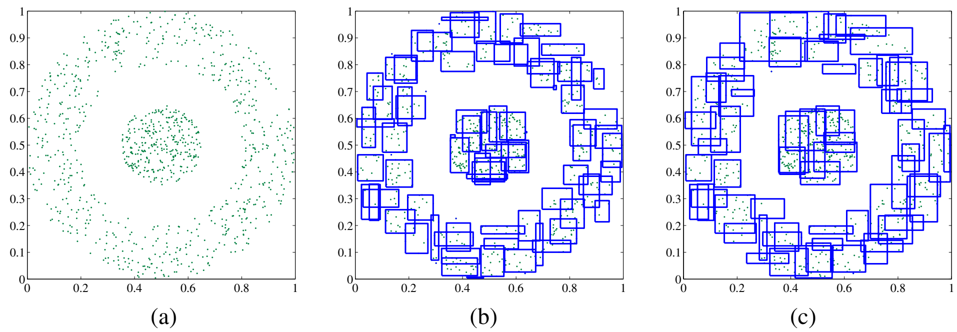


Fig. 6. Simulation of FART and C-FART for circle input data. (a) Input data. (b) Simulation result of conventional FART. (c) Simulation result of proposed C-FART.

TABLE II

NUMBER OF CATEGORIES IN SIMULATION 2.

Algorithm	Number of categories		
	Minimum	Maximum	Average
FART	57	70	62.77
C-FART	48	63	54.62
Improvement rate	10.00%	15.79%	12.98%

TABLE III

NUMBER OF CATEGORIES IN SIMULATION 3.

Algorithm	Number of categories		
	Minimum	Maximum	Average
FART	75	92	82.09
C-FART	60	85	70.34
Improvement rate	7.61%	20.00%	14.31%

ring-shaped cluster contain 300 and 700 points, respectively. The learning conditions are the same as Simulation 1.

The learning result of the conventional FART and the proposed C-FART are shown in Fig. 6(b) and (c), respectively. On the whole, the category size of C-FART is larger than FART because C-FART combines the overlapping categories.

We repeat this learning 100 times and examine the minimum, maximum and average values of the number of categories. The input data are sorted by random each time. The results are summarized in Table III. All the number of categories of the proposed C-FART are smaller than the conventional FART, and C-FART reduces categories by an average of 14.31%. From these results, we can say that the proposed C-FART can perform more effective learning than the conventional FART and can reduce the category proliferation problem by combining the overlapping categories with due consideration of their connections.

V. CONCLUSIONS

In this study, we have proposed Fuzzy ART Combining Overlapped Category in consideration of connections (C-FART) to solve the category proliferation problem. C-FART makes connections between similar categories or releases connections at each step, in addition, C-FART combines the overlapping categories using the created connections. In other words, C-FART combines the categories with due consideration of their similarity. We have applied C-FART to various input data set, and the learning behaviors of C-FART have been investigated. Furthermore, in order to appreciate the training performance of the two algorithms, we have performed the learning simulation 100 times. We have confirmed that the proposed C-FART can obtain more effective result than the conventional FART and can reduce the category proliferation problem.

ACKNOWLEDGMENTS

This work was partly supported by Yamaha Music Foundation.

REFERENCES

[1] C.A. Carpenter, "Distributed learning, recognition, and prediction by ART and ARTMAP neural networks," *Neural Networks*, vol. 10, pp. 1473–1494, 1997.

[2] G.A. Carpenter, S. Grossberg, D.B. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, pp. 759–771, 1991.

[3] G.A. Carpenter, S. Grossberg, and D.B. Rosen, N. Markuzon, J.H. Reynolds, 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, 1991.

[4] T. Frank, K.F. Kraiss and T. Kuhlen, "Competitive analysis of Fuzzy ART and ART-2A network clustering performance," *IEEE Trans. Neural Networks*, vol. 9, no. 3, pp. 544–559, 1998.

[5] H. Isawa, M. Tomita, H. Matsushita and Y. Nishio, "Fuzzy Adaptive Resonance Theory with Group Learning and its Applications," *Proc. of International Symposium on Nonlinear Theory and its Applications*, pp. 292–295, 2007

[6] A.V. Nandedkar, and P.K. Biswas, "Fuzzy Min-Max neural network classifier with compensatory neuron architecture," *IEEE Trans. Neural Networks*, vol. 18, no. 1, pp. 42–54, 2007

[7] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.