JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 30, 713-726 (2014)

DOI:10.1688/JISE.2014.30.3.11

An Improved Learning Rule for Fuzzy ART*

NONG THI HOA[1,2] AND THE DUY BUI[1]
[1]*Human Machine Interaction Laboratory*
University of Engineering and Technology
Vietnam National University
Hanoi, Vietnam
[2]*Thainguyen University of Information Technology and Communication*
Thainguyen, Vietnam

Clustering is an important tool in data mining and knowledge discovery. Fuzzy Adaptive Resonance Theory, a member of unsupervised neural networks, clusters data effectively because of applying operator AND of Fuzzy Logic. In previous studies, learning from data was ineffective when the surface of data is higher than the surface of weight vector of categories. In this paper, we propose an improved learning rule to learn from data better. In the proposed rule, the weights of wining category are decreased to adapt to each input. Each input shows the effect for categories by a learning parameter. The learning parameter is adjusted until the best stable state and performance of clustering results are achieved. We have conducted experiments on 10 benchmark datasets to prove the effectiveness of the proposed rule. The experiment results showed that Fuzzy ART learned with the improved rule (IFART-Improved Fuzzy Adaptive Resonance Theory) performs better than existing models in complex small datasets.

Keywords: fuzzy ART, adaptive resonance theory, clustering, unsupervised neural network, learning rule

1. INTRODUCTION

Clustering is useful for comprehending a large amount of data based on the similarity between data. In studies on clustering, Adaptive Resonance Theory (ART) is shown to be an unsupervised neural network that clustered data effectively. Fuzzy Adaptive Resonance Theory (Fuzzy ART), a member of ART, possesses three advantages. First, Fuzzy ART tests the similarity and learns from data by applying operator AND of Fuzzy Logic. Second, designs of Fuzzy ART are optimized based on many mathematical studies for choosing parameters based on properties of fuzzy set. Then, the category proliferation of Fuzzy ART is solved by normalizing data. Therefore, Fuzzy ART has applied for hundreds of applications such as document clustering [1], classification of multivariate chemical data [2], analyzing gene expression [3], quality control of manufacturing process [4], and classifying with missing data in a wireless sensor network [5].

Models of Fuzzy ART and ART have developed to improve the ability of clustering. Kenaya and Cheok [6] proposed Euclidean ART that employed the Euclidean distance to measure the similarity and the mean of patterns for the weights of categories. Carpenter *et al.* [7] proposed the first Fuzzy ART showing the stable learning's capability of recognition categories. Isawa *et al.* [8] proposed an additional step, "Group Learning", to

Received February 28, 2013; accepted June 15, 2013.
Communicated by Hung-Yu Kao, Tzung-Pei Hong, Takahira Yamaguchi, Yau-Hwang Kuo, and Vincent Shin-Mu Tseng.
* This work was supported by Vietnam's National Foundation for Science and Technology Development (NAFOSTED) under Grant No. 102.02-2011.13.

present connections between similar categories. Then, an improved Fuzzy ART combining overlapped categories was designed based on connections [9]. Yousuf and Murphey [10] provided an algorithm comparing all categories to the input simultaneously and allowing multiple matching categories to be updated. In above studies, learning from data is ineffective when elements of category's weight vector are smaller than elements of the input. It means that categories do not learn from data although data satisfy the given conditions on the similarity.

In this paper, we propose an improved learning rule of Fuzzy ART that learns from data more effectively. In the learning step, each input causes the weight's decrease of the winning category. We use the learning parameter to show the effect of each input to categories. Initially, the learning parameter is set up approximately based on the size of data. Then, the learning parameter is adjusted until clustering results become stable and reach the highest values. Experiments have conducted with 10 benchmark datasets for classifying to prove the effectiveness of the proposed rule. Fuzzy ART learned by our improved rule is consequently compared to existing models including Fuzzy ART, *K*-mean algorithm, Euclidean ART. The results show that IFART clusters more effectively than existing models in complex small datasets.

The rest of the paper is organized as follows. In Section 2, the background is shown. Related works are reviewed in Section 3. In section 4, we present our learning rule and the procedure for finding an optimized value of the learning parameter. Section 5 shows experiments with 10 standard datasets.

2. BACKGROUND

2.1 ART Network

Adaptive Resonance Theory (ART) neural networks [11, 12] are developed to address the problem of stability-plasticity dilemma. The general structure of an ART network is shown in the Fig. 1.

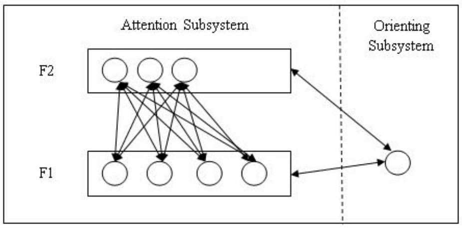

Fig. 1. Architecture of an ART network.

A typical ART network consists of two layers: an input layer (F1) and an output layer (F2). The input layer contains *n* nodes, where *n* is the number of input patterns. The number of nodes in the output layer is decided dynamically. Every node in the output layer has a corresponding weight vector.

The network dynamics are governed by two sub-systems: an attention subsystem and an orienting subsystem. The attention subsystem proposes a winning neuron (or category) and the orienting subsystem decides whether to accept the winning neuron or not. This network is in a resonant state when the orienting subsystem accepts a winning category.

2.2 Fuzzy ART Algorithm

Carpenter *et al.* summarize Fuzzy ART algorithm in [7].

Input vector: Each input I is an M-dimensional vector $I_1, \ldots, I_M$, where each component I_i is in the interval [0, 1].

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

Fuzzy ART algorithm consists of five following steps:

Step 1: Setup weight vector.

Each category j corresponds to a vector $W_j = W_{j1}, \ldots, W_{jM}$ of adaptive weights, or Long Term Memory (LTM) traces. The number of potential categories N ($j = 1, ..., N$) is arbitrary. Initially

$$W_{j1} = \ldots = W_{jM} = 1 \tag{1}$$

and each category is said to be *uncommitted*. After a category is selected to code, it becomes *committed*. As shown below, each LTM trace W_{ji} is monotone non-increasing through time and hence converges to a limit.

Step 2: Choose a wining category.

For each input I and category j, the choice function T_j is defined by

$$T_j(I) = \frac{\|I \wedge W_j\|}{\alpha + \|W_j\|} \tag{2}$$

where the fuzzy AND operator $\wedge$ of Fuzzy Logic is defined by

$$(x \wedge y)_i = \min\{x_i, y_i\} \tag{3}$$

and where the norm $||\cdot||$ is defined by

$$\|x\| = \sum_{k=1}^{n} |x_i|. \tag{4}$$

or notational simplicity, for notational simplicity, $T_j(I)$ in Eq. (2) is often written as T_j when the input I is fixed. The category choice is indexed by j, where

$$T_j = \max\{T_j, j = 1, \ldots, N\}. \tag{5}$$

If more than one T_j is maximal, the category j with the smallest index is chosen.

Step 3: Check state of Fuzzy ART.

Resonance occurs if the match function of the chosen category meets the vigilance criterion; that is, if

$$\frac{\|I \wedge W_j\|}{\|I\|} \geq \rho \tag{6}$$

then learning is performed in Step 4.

Mismatch reset occurs if

$$\frac{\|I \wedge W_j\|}{\|I\|} < \rho \tag{7}$$

then the value of the choice function T_j is reset to -1 for the duration of the input presentation. A new index j is chosen, by Eq. (5). The search process continues until the chosen j satisfies Eq. (6) or actives a new category.

Step 4: Perform learning process.

The weight vector of jth category, W_j is updated according to the following equation:

$$W_j^{new} = \beta(I \wedge W_j^{old}) + (1 - \beta)\, W_j^{old}. \tag{8}$$

Step 5: Active a new category.

For each input I, if no existing category satisfies Eq. (6) then a new category j becomes active. Then, $W_j^{new} = I$.

2.3 Fuzzy ART with Complement Coding

Proliferation of categories is avoided in Fuzzy ART if inputs are normalized; that is, for some $\gamma > 0$

$$\|I\| = \gamma \tag{9}$$

for all inputs I. Normalization can be achieved by complement coding each input vector a.

Complement coding represents both a and the complement of a. The complement of a is denoted by a^c, where

$$a_i^c = 1 - a_i. \tag{10}$$

The complement coded input I to the recognition system is the $2M$-dimensional vector

$$I = (a_i, a_i^c) = (a_1, \ldots, a_M, a_1^c, \ldots, a_M^c). \tag{11}$$

After normalization, $\|I\| = M$ so inputs preprocessed into complement coding form are automatically normalized. Where complement coding is used, the initial condition in Eq. (1) is replaced by

$$W_{j1} = \ldots = W_{j2M} = 1. \tag{12}$$

3. RELATED WORK

Many traditional clustering methods proposed to apply for applications. Three traditional clustering methods are Self Organizing Map (SOM), *K*-mean, and hierarchical clustering. Teuvo Kohonen [13] proposed a model of a new self-organizing process called SOM. SOM is an artificial neural network that performs the unsupervised learning to produce a low-dimensional representation of input space. M. Queen [14] proposed *K*-mean algorithm that classifies a given dataset through a certain number of categories by minimizing the squared error function. Then, the category weight is updated by mean of patterns in each category. Johnson [15] proposed a hierarchical clustering algorithm based on the union between the two nearest categories. The beginning condition is realized by setting each input as a category. After a few iterations, it reaches the final categories. In above studies, the complex of calculating is large because of recalculating the weight vector of every category.

Studies on ART have developed to improve the ability of clustering. A. H. Tan showed a neural architecture termed Adaptive Resonance Associative Map (ARAM) [16]. ARAM extended unsupervised ART systems for rapid, stable, and hetero-associative learning. With maximal vigilance settings, ARAM encoded pattern pairs explicitly and guaranteed perfect storage. Kenaya and Cheok [6] applied the Euclidean neighborhood for the similarity. Then, update weights of a chosen category by the mean of patterns in each category. Lin *et al.* [17] proposed the learning algorithm based on ART for parting online input-output spaces of a traditional fuzzy logic controller. This model established membership functions and found proper fuzzy logic rules based on the distribution of data. The learning process of ART performs ineffectively when the surface of data is higher than the surface of the weight vector of categories.

Studies on theory of Fuzzy ART can be divided into three categories including developing new models, studying properties, and optimizing the performance. In the first category, new models of Fuzzy ART used a general learning rule. Capenter *et al.* [18] proposed Fuzzy ARTMAP for incremental supervised learning of recognition categories and multidimensional maps from arbitrary sequences of input set. This model minimized predictive error and maximized code generalization by increasing the ART vigilance parameter to correct the predictive error. Prediction was improved by training system several times with different sequences of input set, then voting. This vote assigns probability estimations to competing predictions for small, noisy, and incomplete data. Isawa *et al.* [8] proposed an additional step that was called “Group Learning”. An important feature of the learning process was that creating connections between similar categories. It means that this model learned not only weight vectors of categories but also relations among categories in a group. Then, Isawa [9] designed an improved Fuzzy ART combining overlapped categories base on connections. This study arranged the vigilance parameters for categories and varied them in learning process. Moreover, this model voided the category proliferating. Yousuf and Murphey [10] proposed an algorithm that compared the weights of every category with the current input pattern simultaneously and allowed updating multiple matching categories. This model monitored the effects of updating wrong clusters. Weight scaling of categories depended on the "closeness" of the weight vectors to the current input pattern.

In the second category, important properties of Fuzzy ART were studied to choose suitable parameters for each Fuzzy ART. Huang *et al.* [19] presented some vital properties that were distinguished into a number of categories. The vital properties included template, access, reset, and other properties for weight stabilization. Moreover, the effects of the choice parameter and the vigilance parameter on the functionality of Fuzzy ART were presented clearly. Geogiopoulos *et al.* [20] provided a geometrical and clearer understanding of why, and in which order that categories were chosen for various ranges of the choice parameter. This study came in useful when developing properties of learning that pertained to the architecture of neural networks. Anagnostopoulos and Georgiopoulos [21] introduced geometric concepts, namely category regions, in the original framework of Fuzzy ART and Fuzzy ARTMAP. These regions had the same geometrical shape and shared many common and interesting properties. They proved properties of the learning and showed that training and performance phases did not depend on particular choices of the vigilance parameter in one special state of the vigilance-choice parameter space.

In the third category, studies focused on ways to improve the performance of Fuzzy ART. Burwick and Joublin [22] discussed implementations of ART on a non-recursive algorithm to decrease algorithmic complexity of Fuzzy ART. Therefore, the complexity dropped from $O(N*N+M*N)$ down to $O(NM)$ where N was the number of categories and M was the input dimension. Dagher *et al.* [23] introduced an ordering algorithm for Fuzzy ARTMAP that identified a fixed order of training pattern presentation based on the max-min clustering method. The combination of this algorithm with fuzzy ARTMAP established an ordered Fuzzy ARTMAP that exhibited a generalization performance better. Cano *et al.* [24] generated accurate function identifiers for noisy data. This study was supported by theorems that guaranteed the possibility of representing an arbitrary function by fuzzy systems. They proposed two neuron-fuzzy identifiers that offered a dual interpretation as fuzzy logic system or neural network. Moreover, these identifiers can be trained on noisy data without changing the structure of neural networks or data preprocessing. Kobayashi *et al.* [25] proposed a reinforcement learning system that used Fuzzy ART to classify observed information and construct an effective state space. Then, profit sharing was employed as a reinforcement learning method. Furthermore, this system was used to effectively solve partially observed Markov decision process.

4. OUR APPROACH

4.1 Our Improved Learning Rule

The difference between ART and Fuzzy ART, compared by Carpenter *et al.* [7], is that Fuzzy ART uses fuzzy AND operator of Fuzzy Logic for most important steps including choosing a winning category choice, matching the given criteria, and learning.

Reason of applying Fuzzy Logic for ART is that using mathematical studies for optimizing both the design and the performance. Therefore, Fuzzy ART possesses three advantages. First, Fuzzy ART is optimized in the ability of clustering when Fuzzy ART only learns training data that satisfy a given condition. Second, optimized value of parameters is determined based on the geometrical features of data. Then, Fuzzy ART voids the category proliferating. As a result, Fuzzy ART is applied for many applications more and more.

As we discuss in Section 3, ART and Fuzzy ART are ineffective in the learning step. Therefore, we propose an improved learning rule of Fuzzy ART to learn better. In the learning step, the weight vector of categories is dropped for every input. The effect of the input to categories is shown by the learning parameter. We proposed a procedure to find an optimized value of the learning parameter for each dataset. In this procedure, after the learning parameter is setup approximately based on the size of data, it is increased or decreased until clustering results is stable and highest. As a result, IFART can improve the ability of clustering.

Our learning rule updating the weight of the winning category j is presented as follow:

$$W_{ji}^{new} = W_{ji}^{old} - \beta * |I_i - W_{ji}^{old}|, i = 1, \ldots, M \tag{13}$$

where β be the learning parameter.

After updating the weight vector of the winning category, W_{ji} can be modified according to the following rule: If $W_{ji} < 0$ then set $W_{ji} = 0$.

In previous studies, learning rules usually consist of two terms including percents of the old weight vector and remaining percents of the fuzzy AND operator between the input and the old weight vector. For example, Eq. (8) is a typical learning rule of Fuzzy ART. In the proposed learning rule, weights of the winning category also include two terms. The first term is the old weight vector and the second term is the decrease of category's weights. The decrease of category's weights depends on the learning parameter, and the difference between the input and the old weight vector. Therefore, the improved learning rule is different from previous rules of ART and Fuzzy ART.

4.2 Procedure for Finding an Optimized Value of the Learning Parameter

We select a random subset of dataset (about 1/5 records of dataset) which try to possess uniform distribution for categories. Then, IFART uses this subset to test the ability of clustering for each value of the learning parameter.

Step 1: Set up the learning parameter based on the size of dataset based on the following rule: Is bigger the size of dataset, is smaller the learning parameter. Then, calculate clustering results.

Step 2: Do until the clustering result is stable and highest.

- **2.1:** Increase or decrease the value of the learning parameter according to a small step such as 5% of the learning parameter's value.
- **2.2:** Calculate the clustering result.
- **2.3:** Check clustering result:

IF the ability of clustering changes THEN do Step 2

IF the clustering result is stable and highest THEN exit Step 2.
Step 3: Return the optimized value of the learning parameter.

5. EXPERIMENTAL RESULT

We select 10 benchmark datasets from UCI database[1] and Shape database[2] for experiments, namely, Iris, Wine, Jain, Flame, R15, Glass, Pathbased, Aggregation, Blance-Scale, and Spiral. These datasets are different from each other by the number of attributes, categories, patterns, and distribution of categories. Table 1 shows characteristics of selected datasets.

Table 1. Characteristics of datasets.

Index	Name	#Categories	#Attributes	#Patterns
1	Iris	3	4	150
2	Glass	7	9	214
3	Wine	3	13	178
4	Jain	2	2	373
5	Pathbased	3	2	300
6	Spiral	3	2	312
7	R15	15	2	600
8	Flame	2	2	240
9	Aggregation	7	2	788
10	Blance-Scale	3	4	625

Fuzzy ART of Carpenter [7] is implemented into two models including first model (Original Fuzzy ART) and the second model with normalized inputs (Complement Fuzzy ART). Similarly, IFART consists of two models including Original IFART and Complement IFART. We use following models in experiments, namely, Original IFART (OriIFART), Complement IFART (ComIFART), Original Fuzzy ART (OriFART) [7], Complement Fuzzy ART (ComFART) [7], *K*-mean [14], and Euclidean ART (EucART) [6] to prove the effectiveness of IFART.

Data of each datasets are normalized to values in [0, 1]. Initially, we choose a random pattern of each category to be the weight vector. We determine parameters for compared models to reach the highest ability of clustering. For each dataset, we conduct sub-tests with the different number of patterns in each dataset. The percents of successful clustering patterns are presented in a corresponding table. Bold numbers in each table show results of the best model among compared models.

5.1 Experiment 1: Testing with Iris Dataset

Distribution of three categories consists of 1(50), 2(50), and 3(50). Table 2 presents the experimental result with the Iris dataset. Results from Table 2 shows that Complement IFART performs best in all sub-tests.

---
[1] "UCI database," http://archive.ics.uci.edu/ml/datasetss
[2] "Shape database," http://cs.joensuu.fi/sipu/dataset

Table 2. The percents of successful clustering patterns with Iris dataset.

#Record	OriIFART	ComIFART	OriFART	ComFART	EucART	*K*-mean
30	**100**	**100**	**100**	**100**	**100**	**100**
60	98.3	**100**	91.7	**100**	96.7	**100**
90	93.3	**96.7**	72.2	92.2	90.0	94.4
120	95.0	**95.8**	73.3	92.5	90.0	93.3
150	**96.0**	95.3	78.7	92.7	90.0	93.3

5.2 Experiment 2: Testing with Spiral Dataset

Distribution of three categories consists of 1(101), 2(105), and 3(106). Data of Table 3 show that Original IFART is the best in all sub-tests, excepting the last sub-test (smaller 4.1% than the best model – Euclidean ART).

Table 3. The percents of successful clustering patterns with Spiral dataset.

#Record	OriIFART	ComIFART	OriFART	ComFART	EucART	*K*-mean
50	**88.0**	4.0	4.0	22.0	2.0	44.0
100	**71.0**	49.0	37.0	25.0	48.0	22.0
150	**48.0**	33.3	25.3	21.3	32.7	16.7
200	**42.5**	42.0	29.5	39.0	41.0	32.0
250	**40.4**	33.6	27.6	35.2	39.6	32.0
312	38.5	37.8	33.0	28.2	**42.6**	32.7

5.3 Testing with Flame Dataset

Distribution of two categories consist of 1(87) and 2(153). Table 4 shows that Original IFART is the best in all sub-tests, excepting the last sub-test (smaller 3.3% than the best model – Original Fuzzy ART).

Table 4. The percents of successful clustering patterns with Flame dataset.

#Record	OriIFART	ComIFART	OriFART	ComFART	EucART	*K*-mean
50	**100**	**100**	**100**	76.0	88.0	78.0
100	**98.0**	87.0	87.0	83.0	94.0	54.0
150	**98.7**	87.3	80.7	84.7	94.7	69.3
200	**95.0**	76.5	85.5	63.5	74.0	77.0
240	84.6	66.3	**87.9**	55.4	63.3	78.3

5.4 Testing with Blance-Scale Dataset

Distribution of three categories consists of 1(49), 2(288), and 3(288). Results from Table 5 show that Original IFART is greatly better than other models in all sub-tests (5%, 24.5%, 30.33%, 22.25%, and 3.2%), excepting the last sub-test (smaller 5.9% than the second best model – Original Fuzzy ART).

Table 5. The percents of successful clustering patterns with Blance-Scale dataset.

#Records	OriIFART	ComIFART	OriFART	ComFART	EucART	*K*-mean
100	**41**	37	36	30	10	24
200	**69**	44.5	46	42	7.5	25
300	**79.33**	49	46.67	43.33	5	27.67
400	**80**	57.75	49.75	46.25	17	31.25
500	**67**	63.8	57.2	51.2	32.2	28.2
625	53.6	**63.68**	59.52	55.52	45.76	33.6

5.5 Testing with R15 Dataset

Distribution of 15 categories is 40 for each category. Data of Table 6 show that Complement IFART is equal to the best model (Euclidean ART) in four last subtests and lower a bit in two first sub-tests.

Table 6. The percents of successful clustering patterns with R15 dataset.

#Record	OriIFART	ComIFART	OriFART	ComFART	EucART	*K*-mean
100	96.0	98.0	95.0	98.0	**100**	**100**
200	95.5	95.5	93.5	95.5	**96.0**	73.0
300	95.3	**95.7**	88.3	**95.7**	**95.7**	53.7
400	96.0	**96.8**	86.8	**96.8**	**96.8**	64.0
500	96.8	**97.4**	89.4	**97.4**	**97.4**	71.2
600	97.3	**97.8**	91.2	**97.8**	**97.8**	76.0

5.6 Testing with Glass Dataset

Distribution of seven categories consists of 1(70), 2(76), 3(17), 4(0), 5(13), 6(9), and 7(29), especially the distribution of the fourth category is 0. Table 7 shows that Original IFART is significantly greater than the best model in three last sub-tests (15.3%, 13.5%, and 12.6%), and sharply lower in two first sub-tests.

Table 7. The percents of successful clustering patterns with Glass dataset.

#Record	OriIFART	ComIFART	OriFART	ComFART	EucART	*K*-mean
50	16.0	2.0	12.0	12.0	8.0	**82.0**
100	49.0	26.0	42.0	25.0	25.0	**53.0**
150	**56.0**	35.3	36.0	40.7	30.7	35.3
200	**53.0**	43.5	39.5	33.5	36.5	36.5
214	**55.6**	46.3	43.0	37.4	36.9	40.7

5.7 Testing with Wine Dataset

Distribution of three categories consists of 1(59), 2(71), and 3(48). Results from Table 8 show that Complement IFART is approximately equal to the best model (*K*-mean).

Table 8. The percents of successful clustering patterns with Wine dataset.

#Record	OriIFART	ComIFART	OriFART	ComFART	EucART	*K*-mean
30	**100**	**100**	**100**	73.3	**100**	**100**
60	98.3	98.3	98.3	68.3	98.3	**100**
90	83.3	88.9	85.6	64.4	66.7	**90.0**
120	76.7	84.2	82.5	60.8	50.0	**86.7**
150	77.3	**85.3**	82.7	64.7	40.7	**85.3**
178	77.5	**87.6**	83.7	69.7	34.3	**87.6**

5.8 Testing with Jain Dataset

Distribution of two categories consists of 1(276) and 2(97). Data of Table 9 shows that Complement IFART is approximately equal to the best model (Complement Fuzzy ART).

Table 9. The percents of successful clustering patterns with Jain dataset.

#Record	OriIFART	ComIFART	OriFART	ComFART	EucART	*K*-mean
100	99.0	**100**	99.0	**100**	**100**	**100**
200	99.5	**100**	99.5	**100**	57.0	**100**
300	96.3	97.7	69.7	**100**	43.0	**100**
373	94.6	94.4	69.2	**99.7**	47.5	97.9

5.9 Testing with Aggregation Dataset

Distribution of seven categories consists of 1(45), 2(170), 3(102), 4(273), 5(34), 6(130) and 7(34). Table 10 shows that Complement IFART is lower a bit than the best model (Euclidean ART) in previous sub-tests but higher (4.8%) than the best model in the last sub-test.

Table 10. The percents of successful clustering patterns with Aggregation dataset.

#Record	OriIFART	ComIFART	OriFART	ComFART	EucART	*K*-mean
200	98.0	96.5	83.5	83.5	**100**	81.5
400	88.8	91.8	68.3	82.0	**98.0**	66.3
600	83.7	93.0	59.5	84.8	**95.5**	65.2
788	69.2	**78.0**	51.3	68.3	73.2	52.9

5.10 Testing with Pathbased Dataset

Distribution of three categories consists of 1(110), 2(97), and 3(93). Results from Table 11 show that Complement IFART is lower than the best model (*K*-mean).

We sum up results from sub-tests of 10 experiments. Table 12 shows the clustering improvement of IFART compared to the second best model.

Table 11. The percents of successful clustering patterns with Pathbased dataset.

#Record	OriIFART	ComIFART	OriFART	ComFART	EucART	*K*-mean
50	60.0	66.0	54.0	**80.0**	58.0	78.0
100	30.0	33.0	27.0	40.0	29.0	**50.0**
150	36.7	38.0	28.0	34.0	34.0	**60.7**
200	48.0	49.0	21.0	25.5	44.5	**64.0**
250	56.8	59.2	34.0	30.8	55.6	**71.2**
300	64.0	66.0	45.0	42.3	63.0	**76.0**

Table 12. Clustering improvement of IFART compared to the second best model.

Dataset Type	Distribution	#Pattern	#Category	#Attribute	Improvement (%)
1	non-uniform with high level	200-400	2 & 3	2	21-32.6
2	non-uniform with medium level	150	3	9	15.3
3	non-uniform with medium level	200-214	7	9	12.6-13.5
4	non-uniform with high level	500	3	2	9.8
5	non-uniform with high level	788	7	7	4.8
6	non-uniform with low level	120-200	2 & 3	2 & 4	2.5-5.3
7	uniform	90-201	2	3 & 4	1.5-5
8	non-uniform with medium level	250	3	2	0.8

Five first rows of Table 12 show that IFART improves significantly the ability of clustering for complex small datasets. Typically, datasets have features that are non-uniform distribution with high level, small/medium number of categories, small/medium number of attributes, and small/medium number of patterns. However, *K*-mean and Euclid ART sometimes cluster better than IFART for datasets whose density of data is high and boundary of categories is clear.

6. CONCLUSION AND FUTURE WORKS

In this paper, we propose an improved learning rule that learns from data better. In the proposed rule, weights of categories are decreased for each input. The decrease of category's weights depends on the learning parameter. The learning parameter is determined to reach the highest ability of clustering. We have conducted experiments with 10 benchmark datasets to prove the effective of the proposed rule. Experimental results show that IFART is suitable for complex small datasets. However, IFART clusters lower than *K*-mean or Euclidean ART with some datasets. We will improve important steps of IFART to cluster better with all types of dataset in the future.

REFERENCES

1. P. Pardhasaradhi, R. P. Kartheek, and C. Srinivas, "An evolutionary fuzzy ART computation for the document clustering," in *Proceedings of International Conference on Computing and Control Engineering*, 2012.

2. X. H. Song, P. K. Hopke, M. Bruns, D. A. Bossio, and K. M. Scow, "A fuzzy adaptive resonance theory – supervised predictive mapping neural network applied to the clas-sification of multivariate chemical data," *Chemometrics and Intelligent Laboratory Systems*, Vol. 41, 1998, pp. 161-170.
3. S. Tomida, T. Hanai, H. Honda, and T. Kobayashi, "Gene expression analysis using fuzzy ART," *Genome Informatic*, Vol. 12, 2001, pp. 245-246.
4. M. Pacella, Q. Semeraro, and A. Anglani, "Adaptive resonance theory-based neural algorithms for manufacturing process quality control," *International Journal of Production Research*, Vol. 42, 2004, pp. 4581-4607.
5. L. E. Parker, "Classification with missing data in a wireless sensor network," *IEEE SoutheastCon*, 2008, pp. 533-538.
6. R. Kenaya and K. C. Cheok, "Euclidean ART neural networks," in *Proceedings of the World Congress on Engineering and Computer Science*, 2008, pp. 963-968.
7. G. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, Vol. 4, 1991, pp. 759-771.
8. H. Isawa, M. Tomita, H. Matsushita, and Y. Nishio, "Fuzzy adaptive resonance theory with group learning and its applications," in *Proceedings of International Symposium on Nonlinear Theory and its Applications*, 2007, pp. 292-295.
9. H. Isawa, H. Matsushita, and Y. Nishio, "Improved fuzzy adaptive resonance theory combining overlapped category in consideration of connections," in *Proceedings of IEEE Workshop on Nonlinear Circuit Networks*, 2008, pp. 8-11.
10. A. Yousuf and Y. L. Murphey, "A supervised fuzzy adaptive resonance theory with distributed weight update," in *Proceedings of the 7th International Conference on Advances in Neural Networks*, LNCS, No. 6063, 2010, pp. 430-435.
11. S. Grossberg, "Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction and illusions," *Biological Cybernetics*, Vol. 23, 1976, pp. 187-212.
12. S. Grossberg, "How does a brain build a cognitive code," *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control*, Chapter I, Reidel Press, MA, 1980.
13. T. Kohonen, "Self-organizing formation of topologically correct feature maps," *Biological Cybernetics*, Vol. 69, 1982, pp. 59-69.
14. J. B. MacQueen, "Some methods for classification and analysis of multivariate ob-ser-vations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281-297.
15. S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, Vol. 32, 1967, pp. 241-254.
16. A. H. Tan, "Adaptive resonance associative map," *Neural Networks*, Vol. 8, 1995, pp. 437-446.
17. C. Lin, C. Lin, and C. S. G. Lee, "Fuzzy adaptive learning control network with on-line neural learing," *Fuzzy Sets and Systems*, Vol. 71, 1995, pp. 25-45.
18. G. A. Capenter, S. Grossberg, and N. Markuron, "Fuzzy ARTMAP – an adaptive resonance architecture for incremental learning of analog maps," *International Joint Conference on Neural Networks*, Vol. 3, 1992, pp. 309-314.
19. M. Geogiopoulos, H. Fernlund, G. Bebis, and G. Heileman, "Fuzzy ART and fuzzy

ARTMAP-effects of the choice parameter," *Neural Networks*, Vol. 9, 1996, pp. 1541-1559.
20. G. C. Anagnostopoulos and M. Georgiopoulos, "Category regions as new geometrical concepts in fuzzy-ART and fuzzy-ARTMAP," *Neural Networks*, Vol. 15, 2002, pp. 1205-1221.
21. T. Burwick and F. Joublin, "Optimal algorithmic complexity of fuzzy ART," *Neural Processing Letters*, Vol. 7, 1998, pp. 37-41.
22. I. Dagher, M. Georgiopoulos, G. L. Heileman, and G. Bebis, "An ordering algorithm for pattern presentation in fuzzy ARTMAP that tends to improve generalization performance," *IEEE Transactions on Neural Networks*, Vol. 10, 1999, pp. 768-778.
23. M. Cano, Y. Dimitriadis, E. Gomez, and J. Coronado, "Learning from noisy information in FasArt and FasBack neuro-fuzzy systems," *Neural Networks*, Vol. 14, 2001, pp. 407-425.
24. K. Kobayashi, S. Mizuno, T. Kuremoto, and M. Obayashi, "A reinforcement learning system based on state space construction using fuzzy ART," in *Proceedings of SICE Annual Conference*, 2005, pp. 3653-3658.

Nong Thi Hoa received the Master of Science in Information Technology from Thainguyen University in 2006. She is pursuing the Ph.D. degree in Computer Science at University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam. She is now a teacher in Thainguyen University, Vietnam. Her current research interests include artificial intelligence, machine learning.

The Duy Bui received his Bachelor of Computer Science from University of Wollongong in 2000, Australia. He then got his Ph.D. from University of Twente in 2004, the Netherlands. He is now an Associate Professor in Human Machine Interaction Laboratory, which he founded in Vietnam National University. His research interests include embodied intelligent agents, image processing, computer vision and machine learning.