Properties of Learning of a Fuzzy ART Variant

Michael Georgiopoulos*, Issam Dagher*, Gregory L. Heileman**, and George Bebis*

* Department of Electrical and Computer Engineering
University of Central Florida, Orlando, FL 32816, USA
E-Mail: {mng@ece.engr.ucf.edu, ijd@ece.engr.ucf.edu, geb@ece.engr.ucf.edu }

**Department of Electrical and Computer Engineering
University of New Mexico, Albuquerque, NM 87131, USA
E-Mail: {heileman@houdini.eece.unm.edu }

Abstract

This paper discusses one variation of the Fuzzy ART architecture, referred to as *Fuzzy ART Variant*. The Fuzzy ART variant is a Fuzzy ART algorithm, with a very large value for the choice parameter. Based on the geometrical interpretation of templates in Fuzzy ART we present and prove useful properties of learning pertaining to the Fuzzy ART variant. One of these properties of learning establishes an upper bound on the number of list presentations required by the Fuzzy ART variant to learn an arbitrary list of input patterns presented to it. In previously published work, it was shown that the Fuzzy ART variant performs as well as a Fuzzy ART algorithm with more typical values for the choice parameter. Hence, the Fuzzy ART variant is as a good of a clustering machine as the Fuzzy ART algorithm using more typical values of the choice parameter.

1. Introduction

Adaptive resonance theory was developed by Grossberg ([1]), and a list of ART architectures were introduced in the last ten years by Carpenter and Grossberg, as well as by other researchers in the field. A major separation among all of these ART architectures is based on whether the learning applied is unsupervised or supervised. Unsupervised learning is implemented when a collection of input patterns needs to be appropriately clustered in categories, while supervised learning is utilized when a mapping needs to be learned between inputs and corresponding output patterns. A prominent member of the class of unsupervised ART architectures is Fuzzy ART (see [2]), which is capable of clustering arbitrary collections of arbitrarily complex analog input patterns. Our focus in this paper is Fuzzy ART and its associated properties of learning.

In a recent publication ([3]) Georgiopoulos et al., identified three distinct variations of Fuzzy ART architectures depending on the value of the choice parameter α_a: (i) choice parameter small ($\alpha_a \rightarrow 0$), (ii) choice parameter of intermediate value ($0 < \alpha_a < \infty$), and (iii) choice parameter large ($\alpha_a \rightarrow \infty$). This classification was based on the order according to which committed categories are chosen in a Fuzzy ART architecture. In this work, we will focus our attention on one of these variants of Fuzzy ART, the one where the choice parameter is large. From this point on, we will refer to this Fuzzy ART architecture as *Fuzzy ART variant.* Our primary concern with this Fuzzy ART variant is the development of properties of learning, that is properties that help us shed additional light on how learning proceeds in this neural network architecture.

2. Fuzzy ART

2.1 Fuzzy ART Architecture

The Fuzzy ART neural network architecture is shown in Figure 1(a). It consists of two subsystems, *the attentional subsystem*, and the *orienting subsystem.* The attentional subsystem consists of two fields of nodes denoted F_1^a and F_2^a. The F_1^a field is called the *input field* because input patterns are applied to it. The F_2^a field is called the *category or class representation field* because it is the field where category representations are formed. These category representations represent the clusters to which the input patterns, presented at the F_1^a field, belong. The orienting subsystem consists of a single node (called the *reset node*), which accepts inputs from the F_1^a field, the F_2^a field (not shown in Figure 1(a)), and the input pattern applied across the F_1^a field. The output of the reset node affects the nodes of the F_2^a field.
Some preprocessing of the input patterns of the pattern clustering task takes place before they are presented to Fuzzy ART. The first preprocessing stage takes as in-

0-7803-4122-8/97 $10.00©1997 IEEE

put an M_a-dimensional input pattern from the pattern clustering task and transforms it into an output vector $\mathbf{a} = (a_1, \ldots, a_{M_a})$, whose every component lies in the interval [0, 1] (i.e., $0 \leq a_i \leq 1$ for $1 \leq i \leq M_a$). The second preprocessing stage accepts as an input the output a of the first preprocessing stage and produces an output vector $\mathbf{I}$, such that

$$\mathbf{I} = (\mathbf{a}, \mathbf{a}^c) = (a_1, \ldots, a_{M_a}, a_1^c, \ldots, a_{M_a}^c) \quad (1)$$

where

$$a_i^c = 1 - a_i \quad ; \quad 1 \leq i \leq M_a. \quad (2)$$

The above transformation is called *complement coding.* The complement coding operation is performed in Fuzzy ART at a preprocessor field designated by F_0^a (see Figure 1(a)). From now on, we will refer to the vector $\mathbf{I}$ as the *input pattern.*

We denote a node in the F_1^a field by the index i ($i \in \{1, 2, \ldots, 2M_a\}$), and a node in the F_2^a field by the index j ($j \in \{1, 2, \ldots, N_a\}$). Every node i in the F_1^a field is connected via a bottom-up weight with every node j in the F_2^a field; this weight is denoted by W_{ij}^a. Also, every node j in the F_2^a field is connected via a top-down weight with every node i in the F_1^a field; this weight is denoted by w_{ji}^a. The vector whose components are equal to the top-down weights emanating from node j in the F_2^a field is designated by $\mathbf{w}_j^a$ and it is called a *template.* Note that $\mathbf{w}_j^a = (w_{j1}^a, w_{j2}^a, \ldots, w_{j,2M_a}^a)$ for $j = 1, \ldots, N_a$. The vector of bottom-up weights converging to a node j in the F_2^a field is designated by $\mathbf{W}_j^a$. Note that $\mathbf{W}_j^a = (W_{1j}^a, W_{2j}^a, \ldots, W_{2M_a,j}^a)$ for $j = 1, \ldots, N_a$. Initial values of the bottom-up and top-down weights are designated by $W_{ij}^a(0)$, and $w_{ji}^a(0)$, respectively. Initial values of the top-down weights are chosen equal to one. Initial values of the bottom-up weights are chosen equal to:

$$\frac{1}{\alpha_a + M_u^a} \quad (3)$$

where α_a and M_u^a are Fuzzy ART parameters. The parameter α_a is called the *choice parameter* and it takes values in the interval $(0, \infty)$. The parameter M_u^a takes values in the interval $[2M_a, \infty)$; we name this parameter the *uncommitted node choice parameter.* The initial bottom-up and top-down weight choices in Fuzzy ART correspond to the values of these weights prior to presentation of any input pattern to the Fuzzy ART architecture.

In the original Fuzzy ART paper ([2]), only the top-down weights of the architecture are introduced. We have followed a different approach in this paper, intoducing both bottom-up and top-down weights, so that we can naturally define the uncommitted node choice parameter M_u^a which plays a significant role in the introduction of the Fuzzy ART variant in Section 3.

At this point it is important to introduce the notation $\mathbf{w}_j^{a,old}$, $\mathbf{W}_j^{a,old}$, $\mathbf{w}_j^{a,new}$ and $\mathbf{W}_j^{a,new}$. Quite often, templates and bottom-up weights in Fuzzy ART are discussed with respect to an input pattern $\mathbf{I}$ presented at the F_1^a field of Fuzzy ART. In particular, the notation $\mathbf{w}_j^{a,old}$ or $\mathbf{W}_j^{a,old}$ denotes the template of node j or the bottom-up weight converging to node j in the F_2^a field of Fuzzy ART, *prior* to the presentation of an input pattern $\mathbf{I}$ at the F_1^a field (i.e., before learning due to this pattern presentation is initiated). Furthermore, the notation $\mathbf{w}_j^{a,new}$ or $\mathbf{W}_j^{a,new}$ denotes the template of node j or the bottom-up weight converging to node j in the F_2^a field of Fuzzy ART, *after* the presentation of an input pattern $\mathbf{I}$ at the F_1^a field (i.e., after learning due to this input pattern presentation is completed). Similarly, any other quantities defined with a superscript $\{a, old\}$ or $\{a, new\}$ will indicate values of these quantities prior to and after a pattern presentation to Fuzzy ART, respectively.

2.2 Templates in Fuzzy ART: A Geometrical Interpretation

We previously referred to the top-down weights emanating from a node in the F_2^a field as a *template.* A template corresponding to a committed node is called a *committed template*, while a template corresponding to an uncommitted node is called *uncommitted template.* As we have already mentioned, an uncommitted template has all of its components equal to one.

In the original Fuzzy ART paper it is demonstrated that a committed template $\mathbf{w}_j^a$, which has coded input patterns $\mathbf{I}^1 = (\mathbf{a}(1), \mathbf{a}^c(1))$, $\mathbf{I}^2 = (\mathbf{a}(2), \mathbf{a}^c(2))$, $\ldots$, $\mathbf{I}^P = (\mathbf{a}(P), \mathbf{a}^c(P))$, can be written as follows:

$$\mathbf{w}_j^a = \mathbf{I}^1 \wedge \mathbf{I}^2 \wedge \ldots \wedge \mathbf{I}^P = (\wedge_{i=1}^P \mathbf{a}(i), \wedge_{i=1}^P \mathbf{a}^c(i)) \quad (4)$$

or

$$\mathbf{w}_j^a = (\wedge_{i=1}^P \mathbf{a}(i), \{\vee_{i=1}^P \mathbf{a}(i)\}^c) \quad (5)$$

or

$$\mathbf{w}_j^a = (\mathbf{u}_j^a, \{\mathbf{v}_j^a\}^c) \quad (6)$$

where

$$\mathbf{u}_j^a = \wedge_{i=1}^P \mathbf{a}(i) \quad (7)$$

and

$$\mathbf{v}_j^a = \vee_{i=1}^P \mathbf{a}(i). \quad (8)$$

The operations $\wedge$ and $\vee$ stand for the fuzzy min and the fuzzy max operations, respectively. The fuzzy min (max) operation of two vectors produces a vector with components the minimum (maximum) of the corresponding components of these vectors. Based on the above expression for $\mathbf{w}_j^a$, we can now state that the weight vector $\mathbf{w}_j^a$ can be expressed in terms of the two M_a-dimensional vectors $\mathbf{u}_j^a$ and $\mathbf{v}_j^a$. Hence, the weight vector $\mathbf{w}_j^a$ can be represented, geometrically, in terms of two points in the

M_a-dimensional space, $\mathbf{u}_j^a$ and $\mathbf{v}_j^a$. Another way of looking at this is that $\mathbf{w}_j^a$ can be represented, geometrically, in terms of a hyperrectangle R_j^a with endpoints $\mathbf{u}_j^a$ and $\mathbf{v}_j^a$ (see Figure 1(b) for an illustration of this when $M_a = 2$). For simplicity, in this paper we refer to *hyperrectangles* as *rectangles* because all of our illustrations are in the 2-dimensional space.

3. The Fuzzy ART Variant Algorithm

As we have emphasized in the Introduction, our primary focus in this paper is the Fuzzy ART algorithm with a very large choice parameter value α_a (i.e., $\alpha_a \to \infty$). One might question this choice, since when α_a is large Fuzzy ART has the tendency to choose uncommitted nodes over existing committed nodes. This way we may end up with a Fuzzy ART algorithm that does not perform useful clustering since every input pattern from the training list forms its own cluster. This is indeed the case if we assume that the ucommitted node choice parameter M_u^a is chosen equal to $2M_a$, as in the original Fuzzy ART paper ([2]). On the other hand, if we consider a Fuzzy ART architecture with M_u^a very large i.e., $M_u^a \to \infty$, so that committed nodes are chosen prior to any uncommitted node, then it is reasonable to allow α_a to increase to large values, as well. This Fuzzy ART architecture, referred to as *Fuzzy ART variant* in this paper, chooses committed nodes prior to uncommitted nodes, and the criterion for choosing among committed nodes is still the maximum bottom-up input criterion. But now, since α_a is large, the bottom-up input to a committed node j in F_2^a, with template $\mathbf{w}_j^{a,old}$, is proportional to:

$$|\mathbf{I} \wedge \mathbf{w}_j^{a,old}| \qquad (9)$$

where $\mathbf{I}$ stands for the input pattern applied across the nodes of the F_1^a field of Fuzzy ART.
Hence, in review, the Fuzzy ART variant algorithm is the same as the Fuzzy ART algorithm with the following modifications: (i) uncommitted nodes are chosen prior to committed nodes in the Fuzzy ART variant algorithm, and (ii) bottom-up inputs to committed nodes in the Fuzzy ART variant architecture are computed according to equation (8).

4. Properties of Learning of the Fuzzy ART Variant

In this section we report three properties of learning of the Fuzzy ART variant. Then we comment on their importance, and we briefly discuss the basis for their proof. We refer to these properties of learning as Results 1–3. The following results refer to the *off-line training* operation of the Fuzzy ART variant. In the off-line training operation of the Fuzzy ART variant a collection of input patterns is presented to the architecture repeatedly, until learning is over, that is until the weights in the architecture stop changing.

4.1 Statement of Results

Result 1:
Consider the off-line training of a list of P input patterns using the Fuzzy ART variant algorithm. Assume that after the first list presentation the Fuzzy ART variant has created C categories in F_2^a. Designate by $j_i(t)$ $(1 \le i \le C;\ \ t \ge 1)$ the identity of the node with the i-th largest rectangle immediately after the end of the t-th list presentation. Then,

$$\begin{array}{rcl} j_i(t) & = & j_i(i) \\ |R^a_{j_i(t)}| & = & |R^a_{j_i(i)}| \end{array} \qquad \text{for} \quad 1 \le i \le C\,;\ t \ge i+1 \quad (10)$$

Result 2:
Consider the off-line training of a list of input patterns using the Fuzzy ART variant algorithm. Assume that after the first list presentation the Fuzzy ART variant algorithm has created C categories in F_2^a. Designate by $j_i(t)$ $(1 \le i \le C;\quad t \ge 1)$ the identity of the node with i-th largest rectangle immediately after the end of the t-th list presentation. Let S_i $(1 \le i)$ denote the set of training patterns that choose and are coded by node $j_i(i)$ in the $(i{+}1)$-th list presenattion. Then, the patterns of collection S_i will always be coded by node $j_i(i)$ in list presentations $\ge i+2$.

Result 3:
Consider the off-line training of a list of input patterns using the Fuzzy ART variant algorithm. Assume that after the first list presentation the Fuzzy ART algorithm has created C categories in F_2^a. Then, training will be over in at most C list presentations.

4.2 Comments about the Results

Result 1 tells us that the identity of the node with the largest size rectangle does not change after the first list presentation, the identity of the node with the second largest rectangle does not change after the second list presentation, and so on. Result 1, also tells us that the size of the largest rectangle does not change after the first list presentation, the size of the second largest rectangle does not change after the second list presentation, and so on.
Result 2, tells us that patterns that are coded by the largest rectangle in the second list presentation do not need to be presented to the Fuzzy ART variant again, patterns that are coded by the second largest rectangle in the third list presentation do not need to be presented to the Fuzzy ART variant again, and so on. Result 2 is useful because it allows us to eliminate patterns from the training list that do not affect the learning process. This way the learning process can be made to be less computationally intensive.

Result 3 is important because it predicts an upper bound for the number of list presentations required by the Fuzzy ART variant to learn a list of input patterns. In order to identify this upper bound it suffices to present once the collections of input patterns through the Fuzzy ART network (this way we can find the value for the parameter C).

4.3 Proof of the Results

The proofs of the results are based on earlier findings, demonstrated in [4], and presented below as Lemma 1.

Lemma 1:

(a): *In the Fuzzy ART variant, if an input pattern* $\mathbf{I}$ *is inside rectangles* $R_{j_1}^{a,old}$ *and* $R_{j_2}^{a,old}$, *then pattern* $\mathbf{I}$ *will choose the rectangle of the smallest size.*

(b): *In the Fuzzy ART variant, if an input pattern* $\mathbf{I}$ *is inside a rectangle* $R_{j_1}^{a,old}$ *and outside another rectangle* $R_{j_2}^{a,old}$, *then pattern* $\mathbf{I}$ *will choose first rectangle* $R_{j_2}^{a,old}$ *if and only if*

$$|R_{j_2}^{a,new}| < |R_{j_1}^{a,old}| \tag{11}$$

(c): *In the Fuzzy ART variant, if an input pattern* $\mathbf{I}$ *is outside rectangles* $R_{j_1}^{a,old}$ *and* $R_{j_2}^{a,old}$, *then pattern* $\mathbf{I}$ *will choose first rectangle* $R_{j_2}^{a,old}$ *if and only if*

$$|R_{j_2}^{a,new}| < |R_{j_1}^{a,new}| \tag{12}$$

The proofs of the results are based on parts (a) and (b) of Lemma 1. Part (a) and (b) of Lemma 1 are pictorially illustrated in Figures 2(a) and 2(b). The complete proof of the results can be found in [4].

5. Summary–Discussion

In this paper we focused our attention on a Fuzzy ART variant that is obtained if we use very large values for the choice parameter α_a, and the uncommitted node choice parameter M_u^a ($M_u^a >> \alpha_a \to \infty$). Using the geometrical interpretation of the templates in Fuzzy ART we developed useful properties of learning for the Fuzzy ART variant. One of these properties of learning gave us an upper bound on the number of list presentations required by this Fuzzy ART variant to learn an arbitrary list of analog input patterns. This upper bound verified one of the important properties of the Fuzzy ART variant, the fact that off-line training converges to a solution in finite time.

There are two other Fuzzy ART variants that we mentioned in the Introduction of this paper. The Fuzzy ART variant with α_a small ($\alpha_a \to 0$), and the Fuzzy ART variant with α_a assuming intermediate values ($0 < \alpha_a < \infty$). A natural question to ask is whether the results reported in the previous section are valid for these Fuzzy ART variants as well. In the sequel, we elaborate on this topic.

For the Fuzzy ART variant with α_a small, stronger results exist than the ones reported in Section 4. In particular, it was proven in [2] that when α_a is small, Fuzzy ART converges to a solution in one list presentation. Hence, in this case, the identities and sizes of all the rectangles created in the first list presentation remain intact in subsequent list presentations.

For the Fuzzy ART variant with an intermediate α_a value, the results of the previous section cannot be extended in a trivial fashion. The major problem is that Lemma 1 is not valid in its entirety if the α_a value is of intermediate value (see [3]). Hence, we cannot prove Results 1-3 any more, since their validity was demonstrated by using Lemma 1. Proving similar properties of learning, as the ones developed in this paper, for the Fuzzy ART variant with intermediate α_a values is a subject of further research.

It is also worth pointing out that in [4] we evaluated the clustering performance of the Fuzzy ART variant and the Fuzzy ART algorithm for typical values of the choice parameter α_a (e.g., $\alpha_a \leq 10$). The clustering performance of Fuzzy ART and the Fuzzy ART variant was computed for a number of databases included in the UCI repository (see [5]). The conclusion from these evaluations was that the Fuzzy ART variant has performance comparable to the one Fuzzy ART, using typical α_a values, exhibits. Hence, in practice, the Fuzzy ART variant algorithm can be used in clustering applications.

References

[1] S. Grossberg, "Adaptive pattern recognition and universal recoding II: feedback, expectation, olfaction, and illusions," Biological Cybernetics, Vol. 23, pp. 187-202, 1976.

[2] G. A. 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, No. 6, pp. 759-771, 1991.

[3] M. Georgiopoulos, H. Fernlund, G. Bebis, and G. L. Heileman, "Order of search in Fuzzy ART and Fuzzy ARTMAP:A geometrical interpretation," Proceedings of the International Conference on Neural Networks, Washington, DC, pp. 215-220, June 1996.

[4] M. Georgiopoulos, H. Fernlund, G. Bebis, and G. L. Heileman, "Order of search in Fuzzy ART and Fuzzy ARTMAP:A geometrical interpretation," Neural Networks, Vol. 9, No. 9, pp. 1549-1559, 1996.

[5] P. Murphy and D. Ada. UCI Repository of machine learning databases. Technical Report, Department of Computer Science, University of California, Irvine, CA, 1994.

Acknowledgments: This research was supported in part by a grant from Boeing Computer Services under contract W-300445.

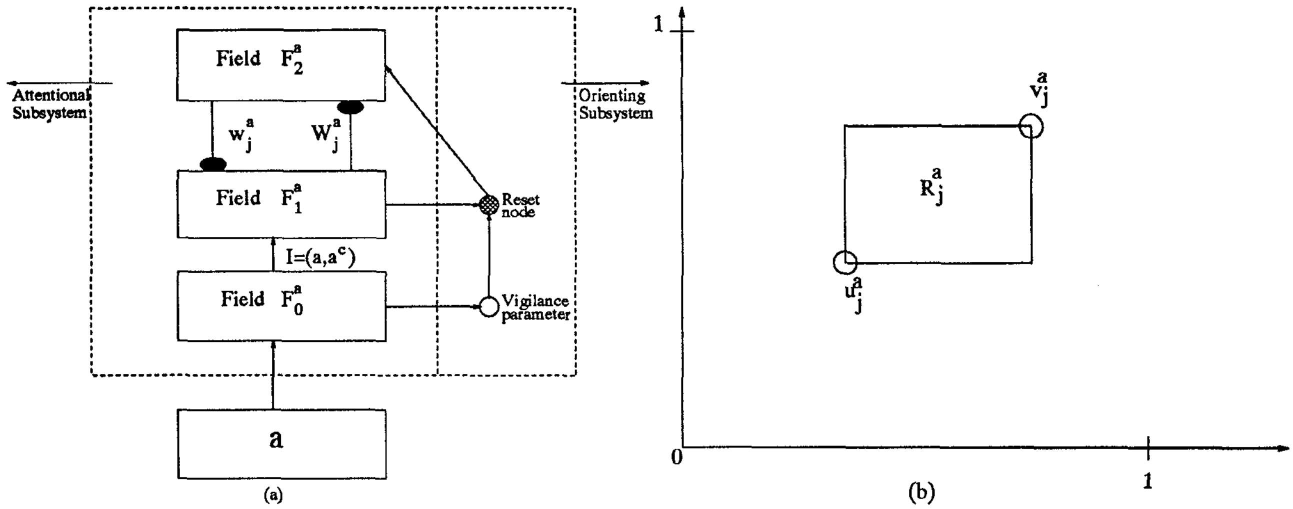

Figure 1: (a): A block diagram of the Fuzzy ART architecture. (b): Representation of the template $w_j^a = (u_j^a, \{v_j^a\}^c)$ in terms of the rectangle R_j^a with endpoints u_j^a and v_j^a.

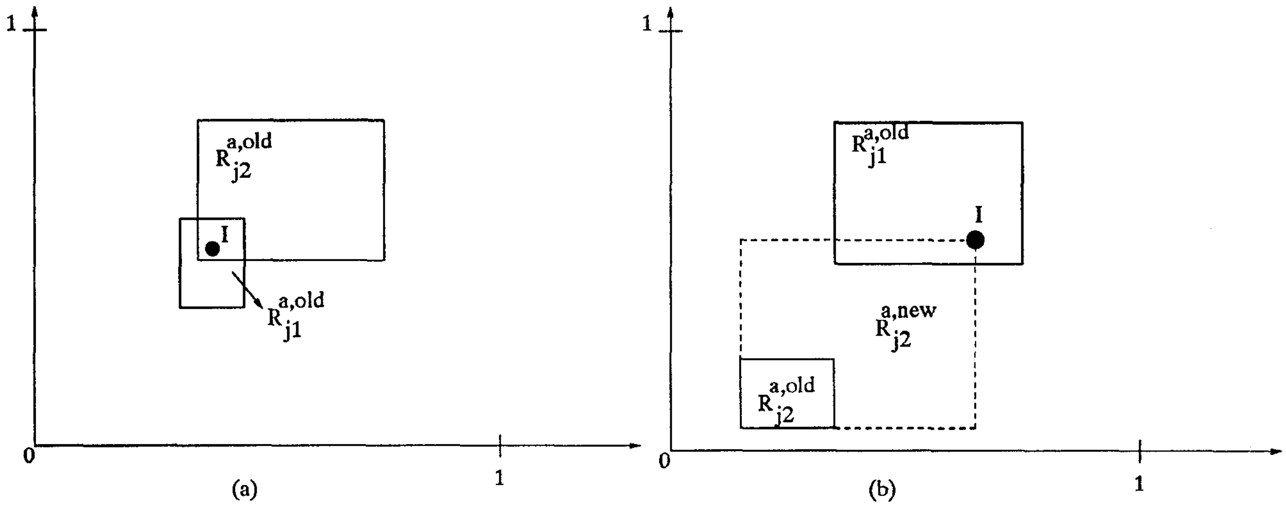

Figure 2: (a): Illustration of Lemma 1: Pattern $I = (a, a^c)$ is inside rectangles $R_{j_1}^{a,old}$ and $R_{j_2}^{a,old}$. Pattern I will choose first the rectangle of the smallest size $R_{j_1}^{a,old}$. (b): Illustration of Lemma 1: Pattern $I = (a, a^c)$ is inside rectangles $R_{j_1}^{a,old}$ and outside rectangle $R_{j_2}^{a,old}$. Pattern I will choose first the rectangle $R_{j_1}^{a,old}$ because $|R_{j_2}^{a,new}| > |R_{j_1}^{a,old}|$.