

Exploring the parameter space of unsupervised ART neural networks for data mining

E. A. Capuano¹ and C. Chauke Nehme²

¹*General Coordination of Computing, Ministry of Social Insurance, Brasilia, Brazil.*

²*Director of Department of Computer Science at Catholic University of Brasilia, Brazil.*

Abstract

Adaptive Resonance Theory (ART), developed by Stephen Grossberg and Gail Carpenter, has attracted some enthusiasts worldwide as a promising approach of artificial neural networks (ANN) for pattern recognition and classification tasks. However, this technology is still not popular among IT professionals and its adoption for data mining applications has not spread out as one could expect. We believe that one reason for this fact is the lack of knowledge about the ART ANN behavior as the parameters change in space. To explore this idea, we executed almost eight hundred experimental tests to map ART1, ART2 and Fuzzy-ART configuration parameters. Some analyses have been done and the results can help system analysts deal with this kind of ANN and explore its complete power. Experimental data, referring to simulated and real databases, have shown that dependent parameter ζ could be clearly defined in Euclidean space R^2 as a function of the vigilance parameter ρ and other patterns and network parameters in ART1, ART2 and Fuzzy-ART models. These results set numerical boundaries up that are useful for data mining application designers using such technology. For instance, the behavior of normalized parameter ζ as a function of ρ shows that it is possible to overcome problems with binary coding in ART1 by imposing higher values of ρ to get complete featured cluster vectors and attenuate (or eliminate) the “bit flood effect” of “0” and “1” observed in “1-of-N” or “thermometer” coding. An important conclusion is that ART1, ART2 and Fuzzy-ART neural networks, as many other ANN approaches, can support data mining applications based on unsupervised learning (for on line systems) overcoming the stability-plasticity dilemma present in other ANN models.

1 Introduction

The discovery of knowledge comes as the most desirable final product of the computation and Fayyad [1] argues that that motivation is outdated, in importance, only by those linked to the preservation of our world and our environment. In the scope of solutions of Business Intelligence, the data mining qualifies itself as a multidiscipline field in the system development for acquisition and administration of knowledge starting from data in the informational environment of the organizations and the key for the success of that system type is its learning capacity along the time. However, as technological products based on artificial intelligence (AI) concepts, systems for data mining aren't still popular among professionals of the computation due to historical cases of failures caused, according to Dhar and Stein [2], mainly for the expertise shortage to transform data in useful knowledge, justifying so, maybe, the need of the creation of a new class of professionals, denominated by Schreiber *et al* "knowledge engineers" [3], capable to build a bridge between the academic world and the real one.

This paper presents experimental results with neural networks based in the Adaptive Resonance Theory (ART), of Carpenter and Grossberg (see references [4], [5], [6], [7] and [8]), a specific model of network that can be used, with advantages, for data mining. The main objective of those experiments is the demystification of the "robotics" of ART in its learning proposal (relatively) stable and flexible, starting from the space mapping of the parameters of configuration of that workmanship of AI, difficulty that comes, usually, in projects of data mining with the use of neural networks. The experimental results, based in about eight hundred tests accomplished with implementations of ART, show that an appropriate configuration of some parameter-keys of the net is crucial for its effectiveness.

2 Engineering of ART nets

The emphasis of Grossberg in the self-learning can be considered as a fourth approach principle to AI based in the nature, that decreed the impossibility of inserting intelligence in a system, one should try to develop that intelligence by interaction of the system with the world. However, Grossberg assumes that the learning that is operated in our brain doesn't leave of a "zero-base", but of a group of preexistent perceptions, looking for support in the secular Theory of the Unconscious Inference, of Helmholtz - in short, the brain tends to notice patterns in a very similar way it hopes to notice them, basing them on previous learning. In terms of architecture of neural networks, the main difference of the ART model with the traditional models is due to the incorporation of pre-conceptions of patterns (in the deductive sense) in the basic algorithm, besides the perceptions of "bottom-up" (inductive sense). The preexistent patterns (of "top-down") represent our cerebral templates, our prejudices in the learning process. In the basic architecture of the ART network only two appealing neural layers

are found, the input patterns or of the memory of short term (layer F1) and the clustering patterns or of the long term memory (layer F2).

2.1 ART1 and Fuzzy-ART

Figure 1 presents the flowchart of the ART algorithm for operations with binary patterns, where z_{ji}^{td} are the synapses (or weights) of the templates (preexistent patterns), I are the vectors of the input patterns, z_{ij}^{bu} are the synapses of the input patterns, L is an engineering parameter of the net and N the number of characteristics (attributes) observable of the input patterns (see details of those parameters in [9]). As for the processing of ART1, it should be observed that: (i) in the initialization of the weights z_{ji}^{td} one attributes them the value "1"; (ii) a neuron of the layer F2 is activated (making $Y_j=1$) when he receives the maximum value of the input T_j (then the others are maintained disabled, with $Y_j=0$); and (iii) the surveillance parameter (it varies between "0" and "1"), being previously stipulated by the user of the net, and the similarity verified between an input pattern and a template should be larger or equal to the expected similarity so that the learning happens (updating of weights).

The appropriate configuration of the parameter ρ is from crucial importance when projecting an ART1 net, because the degree of generalization of the patterns (and of the knowledge) learned by the net depends on it: values of ρ closer to "0" result in more generic clustering, while values closer to "1" result more specific and granular clustering. In practice, the meaning of that domain is richer and the results of the ART1 net with variations of that parameter are presented further on in this paper.

The topology of Fuzzy-ART is identical to the one of ART1, but with Fuzzy-ART there is no distinction among the ascending and descending weights, simply referring to both as z_{ij} . Another aspect that differentiates it is that the operator $\cap$ of ART1 is substituted by the fuzzy operator MIN(or $\wedge$), equivalent to the analogical operator AND. The choice function, in the Fuzzy-ART variant, is given by T_j according to Equation 1, where α is an internal parameter of the net set close to 0 (zero) to avoid that $T_j=1$ happens.

$$T_j = \frac{\|I \wedge Z_j\|}{\alpha + \|Z_j\|} \quad (1)$$

The boundary condition for the surveillance parameter is:

$$\rho \leq \frac{|I \wedge zJ|}{I} \quad (2)$$

The experiments mentioned in this paper were obtained with an implementation proposed by Serrano-Gotarredona *et al* in [9]. In this case, inputs normalized with the Rule of Complemental Code, were used, that consists of the

expansion of the N-dimensional input vectors for input vectors normalized 2N-dimensional, being eliminated, with that, the tendency of proliferation of clustering when operating with small $|I|$ vectors.

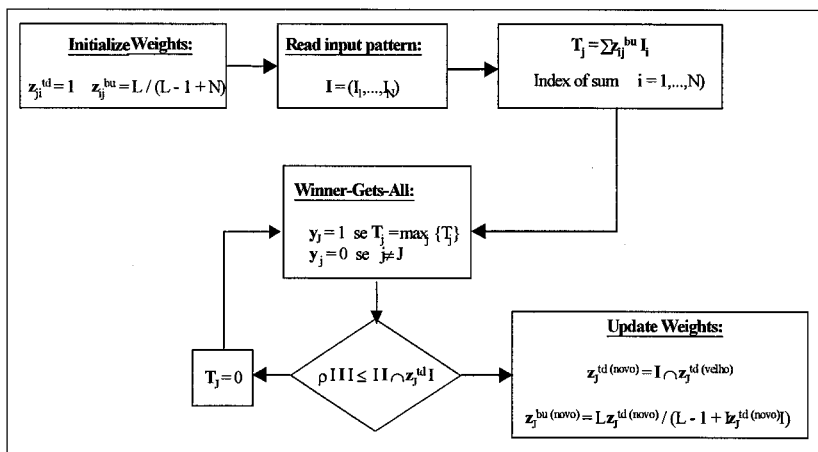


Figure 1: Learning algorithm of ART1 neural networks.

2.2 ART2

This variant, of more generic applications than ART1, is capable to process as binary patterns as analogical. Carpenter and Grossberg [5] propose, in the architecture of ART2, a system of representation of categories (layer F1) for analogical inputs capable to extract and to enhance similar signs encapsulated in contexts with noises that doesn't exist in ART1. The input patterns that flow in the bass sense "bottom-up" and the signs "top-down" are received at different places in the layer F1 and the bows of positive feedback inside of F1 enhance the most salient characteristics of the patterns and suppress the noises. In the original article of Carpenter and Grossberg the best methods of choice of sign functions and ART2 net parameters for categorization of peculiar classes of analogical input patterns for specialized applications are discussed. Table 1 presents the processing equations and the network parameters in the layer F1, revealing normalization mechanisms, earnings control, beat and learning to generate peculiar combinations of emerging desirable properties. Freeman and Skapura [10] present a processing example in the layer F1 being emphasized the importance of the parameter θ as threshold of width of sign applied to the input patterns.

The ART2 variant presents some differentiated aspects that should especially be considered in a project of data mining: (a) as the patterns are normalized before the learning of the network, two input patterns whose values of the characteristics of one are multiples of the values of the characteristics of the other are presented, after the sub layer v of F1, as if they were identical; this

characteristic is useful when operating with analogical values; (b) the characteristics of the learned clustering don't represent the absolute width of their values, but just relative widths amongst themselves, not being, for that, self-explained in knowledge terms; the spectrum of signs of the clustering produced by a ART2 net demands contextual interpretation for recognition of patterns; and (c) in layer F1, the process of search of coincidences of patterns happens in two directions: with the vectors **u** ascending and **p** descending, if both are linearly dependent the search process is, then, interrupted and the input pattern is classified.

Table 1. Processing in the sub layers of ART2.

LAYER	PARAMETERS			
	A	D	J_i^+	J_i^-
W	1	1	$I_i + au_i$	0
x	e	1	w_i	$\ w\ $
u	e	1	v_i	$\ v\ $
v	1	1	$f(x_i) + b.f(q_i)$	0
p	1	1	$u_i + \sum_j g(y_j)Z_{ij}$	0
q	e	1	p_i	$\ p\ $
r	e	1	$u_i + cp_i$	$ u + cp $

As well as in ART1, a lack of coincidences among the **u** vector of ascending input and the descending prototype **p** causes a restoration, when then the network creates a prototype of the input pattern. The calibration of the level of similarity between both patterns **u** and **p**, in ART2, also links with the parameter ρ and it was also object of experimentation, whose results are presented in this paper. Equation 3 presents the parameter **r** as threshold of ρ and as the numeric expression of the level of similarity among the vectors **u** and **p** ("0" means "non similar" and "1" means "totally similar"), where:

$$r = \frac{u + cp}{\|u\| + c\|p\|} \quad \rho > \|r\| \quad (3)$$

$$\|r\| = \frac{[1 + 2\|cp\|\cos(u, p) + \|cp\|]^{\frac{1}{2}}}{1 + \|cp\|} \quad (4)$$

Obviously, in this point of the analysis of the engineering of ART2 the most relevant question as for the configuration of ρ would be on how $\|r\|$ varies (Equation 3). Carpenter and Grossberg show that the minimum point of the function $\|r\|=f(\|cp\|, \cos(u, p))$, in Equation 4, happens with $\|cp\|=1$, and maximal when $\|cp\|=0$ and $\|cp\| \rightarrow \infty$. The maximum of the function of $\|r\|$ reaches the

value of the unit ($\|r\|=1$) and the minimum is close to 0.70 (or $1/\sqrt{2}$). The close points to the minimum of the function are intriguing in a situation of data mining in the real world, because $\|cp\|\approx 0$ is a singularity of occurrence a little probable in the measure in that it demands linear independence perpendicularity among the vectors u and p . The coincidence of both situations would be an event still rarer, being more reasonable to expect values of $\|cp\| \neq 1$ and $\cos(u,p) > 0$. In data bases where the characteristics of the patterns never assume values null or negative, the singularity $\cos(u,p) = 0$ is not common. However, with $\cos(u,p)$ values closer to 0 (zero) the variation of $\|r\|$ has larger width, meaning that it can attribute smaller values of ρ to obtain $\rho > \|r\|$. In practice, this means that ART2 is more efficient when it operates with patterns in no-linear spaces, being little useful in informational atmospheres where the limits among the spaces of patterns tend to the linearity. As can be observed in the results of the experiments presented further on, in spite of the largest number of parameters to be configured in an ART2 network, most of them has width of practical domain reduced, being the effective "calibration" due to the parameters ρ and θ (threshold of sign width).

3 Experiments with configuration parameters

The computational applications used are based in the contributions of Serrano-Gotarredona *et al* [9] and of Chauke-Nehme [11], with the programming tool MATLAB (version 5.0). The environment of experimental processing was composed of microcomputers type IBM PC, Intel pattern. As for the tested patterns, they are composed of simulated data, in most of the cases, and some extracted real datasets of legated systems of transactions [12]. Altogether, 40 files were created with patterns of input couples, as masses of data for tests of ART1, and analogical, for tests of ART2 and Fuzzy-ART. The experiments consumed approximately 400 machine hours along ten months of work. It was observed, in the experiments, the behavior of the main parameters of configuration of the net variants ART1, ART2 and Fuzzy-ART, emphasizing that, for mapping of configuration domains in the Euclidian space, some parameters resulted of the normalization of more elementary parameters as number of patterns of input of the mass of data (variable "np"), number of characteristics of each pattern in its vectorial form (variable "n"), and number of classes or clustering produced by the net (variable "nc").

3.1 Functions of ρ in ART1 and Fuzzy-ART

Starting from the experiments with ART1, it is proven that the network parameters L and ρ are crucial for an appropriate configuration of the model. As can be observed in Figure 2, the curves of the relationship $\rho \times nc/np$, for each area of domain of the parameter of control np/n and of L , present pattern-forms with few variations. The curves of experimental relationships of parameters suggest that: (i) $\zeta = nc/np$ is non linear function of ρ ; (ii) the curves of the

relationship $\rho \times nc/np$ vary a little in function of np/n for values of L close to 1.00, and when $L=5,00$ the curves begin to alter their form in the interval of $0.10 < \rho < 0.50$; (iii) the reason nc/np is inversely proportional to the reason np/n ; (iii) for ρ close to 0.10, nc/np increases inversely to np/n , in a more sensitive way when $L=5,00$; for ρ close to 0.90, ζ is little sensitive to L ; (iv) when $np/n \geq 6$, the curves of $\rho \times \zeta$ assume a pattern-form practically invariable and of upward concavity; and, as larger np/n is, more accentuated that concavity is. Those results confirm the conclusions of Serrano-Gotarredona *et al* [9] that when the value of ρ decreases the parameter L assumes larger responsibility on the medium number of categories learned by the net. In normal conditions of data mining the relationships np/n produce absolute values of the order of hundreds or thousands, when then the clustering curves will come, probably, stableer, described in parametric spaces better defined (closer of the axis of ρ). As for L , the conclusion is that as larger the value of that parameter poorer the curve of the parametric model will be in terms of exploration of the image of c in ζ in the space R^2 , that is, as larger L is, as bigger ρ will have to be so that the subsystem of orientation of the net ART1 begins to operate the processes of production of new categories.

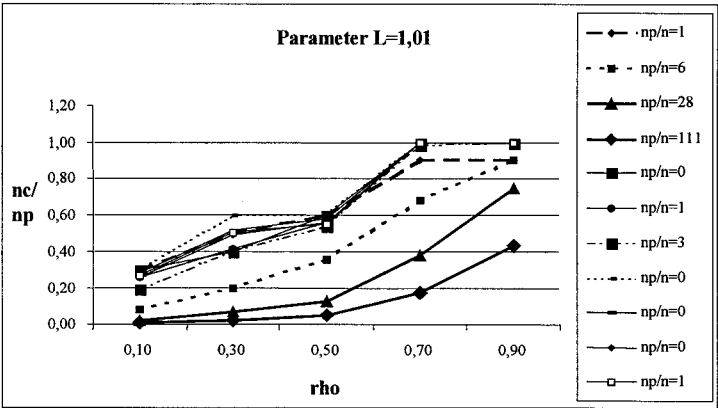


Figure 2: Behavior of parameters with ART1 [12].

The parameters of configuration of Fuzzy-ART don't differ a lot of the parameters of ART1. While in ART1 the parameter L integrates the equations of updating of ascending weights, when restoration of categories happens, in Fuzzy-ART appears the parameter, denominated "choice parameter", with the mission of impeding that all the values T_j may be equal to "1" in case it happens $|I \wedge Z_j| = |Z_j|$ when a pattern of input I is presented to the layer F1. As in ART1, the subsystem of orientation of the Fuzzy-ART network can be activated with lower values of ρ (Figure 3). The curves of the clustering function are very well pronounced and don't suffer significant influence of the parameter α , for

$0.10 \leq \alpha \leq 5.00$. And, even when the relationship of dimensionality np/n produces smaller values than 100, the displacement of the clustering curve upward is not as much as in ART1 and the curve doesn't change its basic topology. As the np/n value increases, for values of $\rho < 0.70$, the clustering curves tend to "crawl", close to the axis of ρ . After that threshold of ρ , when ρ tends to 1 the curves tend to deflect upward, as a parable branch, meaning that the network begins to learn more specific categories of patterns in a more intense rhythm.

The experiments of Capuano [12] showed that the curves of clustering of Fuzzy-ART networks tend to position themselves a little below of the curves of ART1 due to the pertinence of the fuzzy patterns around their attraction areas. The configuration of parameters in the Fuzzy-ART variant doesn't present difficulties. However, for data mining some project recommendations may be useful: (i) to use $\alpha = 0.10$ and values of $\beta = 1$ if the category J has not been tested (slow re-codification); (ii) to estimate the oscillation space of ρ , for beginning of the data mining, in agreement with the clustering function $\zeta = f(np/n, \alpha, \rho)$, changing the value of this parameter upward and downward according to the project needs, as exercise of the engineer's of the knowledge sensitization in relation to the surveillance parameter; in this phase, it is important to discover the value threshold of ρ starting from which the categories begin to proliferate; (iii) for visual observation of the areas of patterns, in R^2 , to use the proposal presented by Serrano-Gotarredona *et al* [9] and, for observations in R^n , to use graphs of bars with two areas in each characteristic (bar) acted.

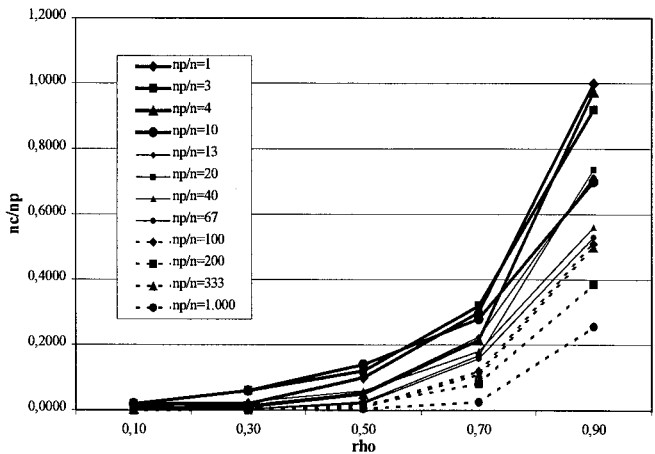


Figure 3: Behavior of parameters with Fuzzy-ART [12].

3.1.2 Functions of ρ in ART2

With ART2, the configuration of the network is a more complex task and it presents some difficulties not experienced in ART1 and Fuzzy-ART. In the first place, the mapping of the spaces of domains of parameters of applications

involves the internal parameter of the patterns $\cos(\mathbf{u}, \mathbf{p})$, as discussed previously. The experiments with the net parameters "a", "b", "c", "d", " θ " and " β ", in ART2, were executed in an area of the domain of ρ where restoration of patterns usually doesn't happen during the learning process, with $\rho < 0.85$. The parameter "e", as proposed by Freeman and Skapura [10], was maintained equal to 0 (zero); it also was set that $a=10$, $b=10$ and $c=0.10$ as independent parameters in the equations of the layer F1 (see Table 1), varying (from 0.10 to close values of 1.00 to each group of input patterns with the defined parameters n and n_p (control variables). The parameter d appears as activation threshold in the function of exit of the layer F2 (sub layer p) and, by suggestion of Freeman and Skapura, in the accomplished experiments was maintained equal to 0.90. In short, those authors suggest the following domains (or boundary conditions) of net parameters ART2:

$$a, b > 0, \quad 0 \leq \theta \leq 1, \quad 0 \leq d \leq 1, \quad 0 \leq \rho \leq 1, \quad cd/(1-d) \leq 1, \quad e \lll 1$$

The parameters "a", "b" and "c" are justified as coefficients of the linear functions of excitatory processing in the sub layers "w", "v" and "r", according to Table 1. With a mass of analogical data of $n_p=1,000$ and $n=5$, with $\rho = 0.80$, we performed experiments for tests of some combinations of "a", "b", "c" and "d", that didn't produce anything new, because it was obtained $n_c=1$ in all of the configuration situations. In the beginning, it was suggested the use of the values proposed by Carpenter and Grossberg [5] for those parameters, being avoided condemnable configurations as: (i) $a = 0$, because, in this case, alteration in the learned patterns doesn't happen, unless one need to enhance the patterns of input I in relation to the weights Z_j in the sense "from top to bottom"; (ii) $b = 0$, because, like this, the system generates an unstable, inconsistent structure of categories, where the input patterns oscillate between a category and other during the presentation sequence; (iii) $d \cong 0$, because then the same phenomenon observed in the case of $a = 0$ happens; (iv) $\theta = 0$, because this way the function $f(x)$ of the layer F1 becomes linear and, consequently, without the mechanism of noise suppression and emphasis of contrasts of the input patterns the network loses the capacity of correct separation of the vectors in categories well differentiated, producing little useful results.

The experiments with variations of θ and ρ were the ones that showed the problems of configuration of the ART2 network ($\beta = 0.25$ was adopted, equivalent to a slow learning tax). The results confirmed the observations of Carpenter and Grossberg [5], about the use of $\theta = 1/\sqrt{n}$, where n is the number of characteristics of the patterns; this is the threshold of sensibility of the ART2 network to not uniform patterns. As for ρ , the experimental results showed an effect completely different from that parameter in ART2, in comparison with ART1 and Fuzzy-ART. In the beginning, it was observed that the network is little sensitive to values of ρ below $1/\sqrt{2}$ (that is the inferior threshold of $\|r\|$). It was observed, in the experiments in general, that ART2 produced clustering just with values of ρ around 0.85 to 0.90 in most of the masses of data tested, meaning that the net produces new categories of clustering only with values of ρ higher

than a characteristic threshold of each mass of input patterns. And it was just with two masses of analogical data of parameters $n_p=100,000$ and $n=9$ (data of registrations of external connections to a local area network) that ART2 produced clustering categories with values of ρ soon starting from $1/\sqrt{2}$, being suggested that those masses of data have, especially, non linear characteristics with the parameter $\|r\|$ close to $1/\sqrt{2}$. However, with ρ below that threshold the ART2 network perform very well the classification of patterns, operating at an appreciable speed, because even with files relatively big in the context the response time of the application was just revealed a subject of seconds.

The experimental curves of $\zeta=f(\rho)$ observed with the implementations of networks ART1, ART2 and Fuzzy-ART can be presented as in the Figure 4, showing a model of quite defined clustering for each variant of the network model.

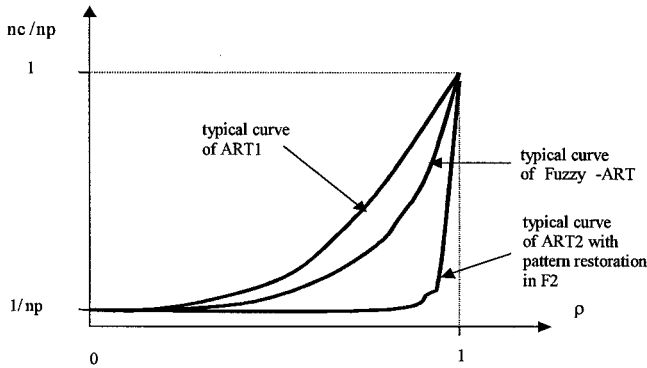


Figure 4: Clustering functions in ART nets [12].

The asymptotic growth of the reason $\zeta = nc/np$ towards the unit happens with values of ρ in a space of intermediate domain (between 0 and 1) in ART1 and Fuzzy-ART, while in ART2 the proliferation of patterns begins to happen only in a domain of ρ very restricted and close to 1,0.

3.2 Bias of the data

The experiments with ART1 [12] in masses of data with bias reinforced the idea that an appropriate configuration of the surveillance parameter ρ is vital not only for trying to establish, *a priori*, the wanted granularity of the clustering of patterns, but also to produce complete patterns in terms of representations of characteristics. The code "1-of-N", for instance, can become uncomfortable and inefficient, under the point of computational view, for relatively high analogical values, because the representation of those values demands, usually, expressive number of zeros for each characteristic of the pattern-vector.

An experiment (see details in [12]) was executed with ART1, being used, as mass of data, a set of register of people, organizing the binary patterns of input with the "thermometer" code in form of a vector of 115 bits visual as if it was a

head office of pixels of 5 x 23 bits. It was observed, starting from the representation of the input patterns, that the resulting binary code "thermometer" characterized each input pattern with bits "1" very concentrated in the characteristic "age" and it caused, for that, the "asphyxiation" of the other binary characteristics in the group of bits of each pattern. The ART1 net, this way, was guided too much by the bits of the characteristic "age" for the recognition of more generic patterns (with values of ρ closer to zero), to the detriment of the other characteristics with smaller amounts than bits "1." An effect observed of that flood of bits "1" with the code "thermometer" (see Table 2), in the executed experiments, was the brutal reduction of the number of clustering produced by the net, in relation to the classic binary code, for a same value of ρ . Like this, complete patterns of clustering, with all the not-null characteristics, were recognized by ART1 only starting from values of ρ closer to 0.50, or even larger.

With the code "thermometer", the most appropriate for ART1 in terms of space of disconnection of patterns, it was also concluded that more satisfactory results of clustering can be obtained, with more complete patterns in terms of represented characteristics, by a configuration of ρ with intermediate values or values closer to 1,00.

Table 2. Formation of clustering in function of the code in ART1 [12].

ρ	PATTERNS OF 115 BITS							PATTERNS OF 54 BITS						
	NUMBER OF CLUSTERS WITH “X” CHARACTERISTICS NON-NILL													
	X=	X=	X=	X=	X=	X=	Tot.	X=	X=	X=	X=	X=	X=	Tot
	1	2	3	4	5	6		1	2	3	4	5	6	
0.10	2	0	0	0	0	0	2	10	7	0	1	0	0	18
0.20	3	0	0	0	0	0	3	8	32	28	5	0	0	73
0.30	5	0	0	0	0	0	5	3	30	96	50	12	0	191
0.40	6	0	1	1	0	0	8	3	42	167	122	35	4	373
0.50	9	1	3	1	1	0	15							
0.60	19	5	5	3	0	1	33							

4 Conclusions

The experiments suggest stability of relationships among internal network parameters, external parameters of input patterns and clustering parameters in ART networks. Based on this experimental knowledge, it is suggested that those parameters can be configured on an appropriate way to each application type for data mining in an informational ecosystem, in a way to turn the computational environment more effective and efficient, with perspective of more economical costs of project. The starting point is the generic curve of clustering of each variant of that model, as in Figure 4, when the knowledge engineer will then be able to deduce the other useful parameters involved in the project of data mining, such as "nc" and "np". Knowing, by approach, the clustering curve, one can

determine the minimum value of the parameter n_p (or minimum volume of the mass of data), making $n_p = n_c / \zeta$, where $\zeta = f(p_i)$, being p_i the group of the internal network parameters, external parameters of the patterns and p . The definition of the parameter “ n_c ” should be a function of the informational environment and of the objectives of the project of data mining in the organization, because the need of larger generalization implicates smaller “ n_c ”, and vice-versa. With the progressive experimental mapping of the spaces of parameters, it is expected that that cognitive paradigm of artificial neural network is shown more and more previsible, reliable, efficient and, therefore, appropriate for use in data mining applications.

References

- [1] Fayyad, Usama M. *et al.* Advances in Knowledge Discovery and Data Mining. Menlo Park (California), AAAI Press (MIT Press), 1996.
- [2] Dhar, Vasant and STEIN, Roger. Seven Methods for Transforming Corporate Data into Business Intelligence. Prentice-Hall, 1997.
- [3] Schreiber, Guus *et al.* KADS: a Principled Approach to Knowledge-Based System Development. London and San Diego, Academic Press, 1993.
- [4] Grossberg, Stephen. Nonlinear Neural Networks: principles, mechanisms, and architectures. In: Neural Networks, vol. 1, 1988.
- [5] Carpenter, Gail A. and GROSSBERG, Stephen. ART 2: self-organization of stable category recognition codes for analog input patterns. In: Applied Optics, vol. 26, No. 23, December 1987.
- [6] Carpenter, Gail A. and GROSSBERG, Stephen. The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network. In: IEEE, March 1988.
- [7] Carpenter, Gail *et al.* Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System. In: Neural Networks, vol. 4, 1991.
- [8] Carpenter, Gail *et al.* Fuzzy ARTMAP: a Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps. In: IEEE Transactions on Neural Networks, vol. 3, No. 5, September 1992.
- [9] Serrano-Gotarredona, Teresa *et al.* Adaptive Resonance Theory Microchips: circuit design techniques. Boston, Dordrecht and London, Kluwer Academic Publishers, 1998.
- [10] Freeman, James A. and SKAPURA, David M. Redes Neuronales: algoritmos, aplicaciones y técnicas de programación. Wilmington, Delaware (EUA), Addison-Wesley Iberoamericana, 1993.
- [11] Chauke-Nehme, Cláudio. Neurocomputacional Model of Off-center Elements with Surrounding Interactions applied to Spacetime Pattern Recognition. Rio de Janeiro, COPPE/UFRJ, Federal University of Rio de Janeiro, 1996. Doctorate Thesis (Original version in Portuguese).
- [12] Capuano, E. A. Data Mining with Unsupervised ART Model Neural Networks: parameters and applications. Brasília, Catholic University of Brasília, 2001. Master Thesis (Original version in Portuguese).