VALIDATING AN UNSUPERVISED WEIGHTLESS PERCEPTRON

Iuri Wickert and Felipe M. G. França

Computer Systems Engineering Program - COPPE, UFRJ
Caixa Postal 68511, 21945-970, Rio de Janeiro, RJ, Brazil
iwickert@yahoo.com, felipe@cos.ufrj.br

ABSTRACT

This paper presents a comparison between two unsupervised neural network models: (i) the well-known Fuzzy ART, and (ii) AUTOWISARD, a new unsupervised version of the classic WISARD weightless neural network model. It is shown that AUTOWISARD is simple, fast and stable, whilst keeping compatibility with the original WISARD architecture. Experimental test results over binary patterns benchmarks have shown that, although both unsupervised learning models are remarkably simple, AUTOWISARD consistently exhibits better classification skills than Fuzzy ART. It is also shown that such superiority happens thanks to AUTOWISARD's richer internal representation of the trained patterns and the training methods employed by the algorithm, such as the *learning window* and *partial training* strategies.

1. INTRODUCTION

The WISARD is a well-known weightless supervised neural network that is able to do fast, one-shot learning of binary patterns and, despite its simplicity, do have good generalisation abilities. The AUTOWISARD model is an unsupervised version of the WISARD, extending it with an unsupervised operation mode similar of the also classic ART model's, but adding nothing to its standard architecture.

Although both the AUTOWISARD and ART are similar to each other in many ways, they are by no means identical: in fact, the AUTOWISARD's class representation mechanism (discriminator) is a generalisation of the ART class vectors in the sense that a single discriminator can represent a whole set of ART vectors. To demonstrate the complex generalisations that can be obtained by an AUTOWISARD network, a simple handwritten digits images classification task was deployed, using the newer, simpler Fuzzy ART as the comparative standard. The experimental results showed that, even when both networks generated classifications with the same number of classes, the AUTOWISARD's consistently showed better classification skill than Fuzzy ART's, consequence of the WISARD's "multivetorial" internal representation, monotonic training and the AUTOWISARD's own training control mechanisms.

The remainder of this paper is organized as follows. Sections 2 and 3 present Fuzzy ART and AUTOWISARD unsupervised neural models, respectively. Section 4 contains the description of the experiments defined for the comparison between the two target neural models. Evaluation results are discussed in Section 5. Section 6 shows the conclusions of this work.

2. FUZZY ART

The Fuzzy ART [4] is a simpler and more recent variation of the classic ART unsupervised neural network [5]. Its main purpose is to solve the "stability-plasticity dilemma": to develop a networks that is able to learn when confronted with new inputs (plasticity) whilst beign able to recognise previously learned inputs (stability).

The Fuzzy ART consists of a set of binary vectors (nodes), each representing a distinct class of patterns. These vectors can be updated (taught) to embody new knowledge, or new vectors can be allocated to contain new classes of inputs. The training procedure of a Fuzzy ART is dependent on two measurements; assuming X_i an input patterns and W_i an network node, the similarity and resonance measures are defined as follows:

$$sim = \frac{\sum_{j=1}^{f} (X_{i,j} \wedge W_{i,j})}{\sum_{j=1}^{f} W_{i,j} + \alpha}, \quad \alpha = 10^{-6}$$

$$res = \frac{\sum_{j=1}^{f} (X_{i,j} \wedge W_{i,j})}{\sum_{j=1}^{f} X_{i,j}} > \rho, \quad 0 \leq \rho \leq 1$$

and the updating of a class vector is defined as:

$$W_i = W_{i,j} \wedge X_{i,j}$$

For a given Fuzzy ART network and an input pattern, a similarity list (a list containing all the classes indices, sorted by the similarity measure) is created, and, from the most similar class on, if that class passes the resonance test (for a generalisation control parameter ρ), the class vector will be updated. In case no class passed that test, a new class vector will be appended to the network, using as initial value a copy of the input pattern. The Fuzzy ART stabilizes its state after a single training epoch; although the presence of a previously trained pattern could trigger the network to update one of its classes again, the resulting class vector will keep the same values.

3. AUTOWISARD

The AUTOWISARD model [1] is an attempt to bring the features of more mainstream unsupervised learning models (automatic allocation of new classes as needed; to solve the "stability-plasticity dilemma") to the supervised weightless model WISARD [2][3], leveraging its original characteristics, namely simplicity, fast training/recognition, monotonic training and powerful class representation.

The WISARD's basic processing element (neuron) is a RAM-type memory unit, having n address inputs and able to store 2^n bits (all positions are initialized with 0's). That way, each neuron is able to learn and recognize n-bit words ("tuples"). The training of a new tuple consists in writing '1' on the neurons' position addressed by it; the positive recognition of an input tuple bu a neuron is merely checking if it addresses a stored '1'.

The RAM neuron, while fast, lacks generalisation power: it only recognises previously learned tuples. To overcome this limitation, a set of neurons can be organized in a structure called discriminator, where each neuron is responsible for the learning/recognition of a subset of a (larger) input pattern (Figure 1). The subpattern assigned to each neuron is defined in a randomly created input-neuron mapping, which is used in both learning and recognition phases. The training of a discriminator consists in writing 1's in the positions addressed in each of the discriminators' neurons. The recognition of a pattern is given by analysing the discriminator's output, which is the sum of the neurons' outputs bits for that pattern. By having a graded output, a discriminator can recognise similar but different version of a trained pattern, thus showing generalisation ability.

The WISARD network is an array of discriminators, each representing a different class of patterns (Figure 2). The WISARD is trained in a supervised fashion: one must select a discriminator and train it with selected patterns from the respective class. The determination of a pattern's class by a WISARD is made in a competitive way: the input pattern belongs to the discriminator which presented the highest recognition level (output) for that input, assigning to itself the winner discriminators' class label.

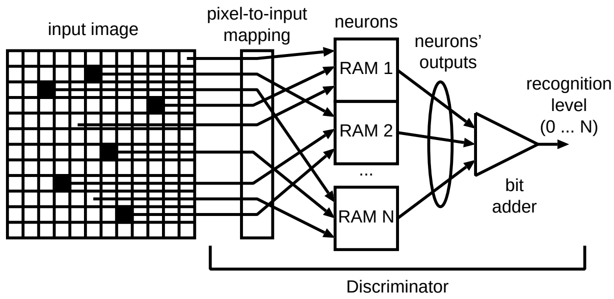

Figure 1: The discriminator.

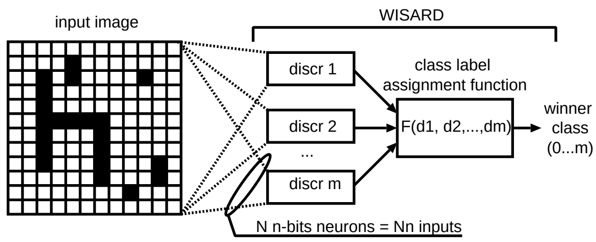

Figure 2: The WISARD network.

The AUTOWISARD model is a WISARD network which trains the winner discriminator with a well-matching input pattern and dynamically creates a new one when the pattern couldn't be acceptably recognised by the existing ones. To decide when to create new discriminators, it implements a *learning window*, which is a region over the discriminators' recognition interval, defined by the parameters w_{min} and w_{max}, $0 \leq w_{min} \leq w_{max} \leq r_{max}$, r_{max} being the maximum recognition value for a discriminator, i.e., the number of RAM neurons (Figure 3). The learning window operates as follows: during the training phase, given r_{best} as the WISARD's best recognition for a given input pattern, if

- $0 \leq r_{best} \leq w_{min}$, a new discriminator is created and trained with that pattern (the pattern isn't represented by the current state of the discriminators);
- $w_{min} < r_{best} < w_{max}$, it can either allocate a new discriminator or submit the winner discriminator to a *partial training*. The action to be performed is probabilistically selected, relative to the distance of r_{best} to w_{max}: a pattern whose recognition is closer to w_{max} have a greater probability of suffering a partial training, if r_{best} is closer to w_{min}, then the probability of allocating a new discriminator is greater (Figure 3);
- $w_{max} \leq r_{best}$, the training algorithm does nothing; it is assumed that the input pattern is already well represented by an existing discriminator.

The partial traning consists in training just enough neurons (from the set of mismatching ones) of the winner discriminator to make $r_{best} = w_{max}$, for the given input pattern. That kind of training ensures that every input pattern which was trained by the network has a recognition value of w_{max}, and also helps to prevent leading the discriminators to a saturation state (excessive number of stored 1's), which harms their discriminative power.

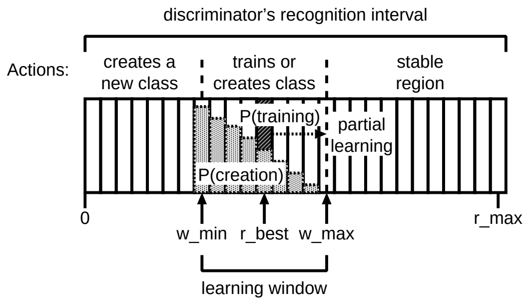

Figure 3: Representation of the AUTOWISARD control methods.

4. EXPERIMENTS

To provide the experimental data for the comparison between the two neural networks, a simple OCR-like application was devised. It was used as input data a set of 1934 labeled, binary images of handwritten digits [6]: each image has a size of 32x32 pixels, and the digits have a reasonably even frequency in the set. Some samples of the digits are shown in the Figure 4.

The experiments consisted in presenting the whole input set, with different random orders, to train, without supervision, each neural network (2 runs per model), then use the same set to label the classes generated by the networks and to extract measurements over that classes. The measurements considered in this work are:

- The number of recognised images for each class
- The label of the most frequent digit in each class
- The frequence of the most frequent digit, relatively to its class' set of images
- The class' average radius in bits, using the Hamming distance measure

and were meant to illustrate the quality of the classifications generated by the 4 training/recognition runs.

As the quality of a classification is dependent of the number of generated classes (regardless of the classification method employed), and so is the degree of generalisation presented by them, for comparison purposes, all the 4 test runs have the same number of classes (20). The control parameters used in the Fuzzy ART runs was the vigilance parameter $\rho = 0.31$, and for the AUTOWISARD were 4 bits (256 neurons per discriminator), learning window parameters $w_{min} = 30\%$ and $w_{max} = 38\%$; all runs using these configurations which didn't generated exactly 20 classes were discarded. This way, the influence of the generalisation factor is minimized, allowing a more faithful comparison of these different networks.

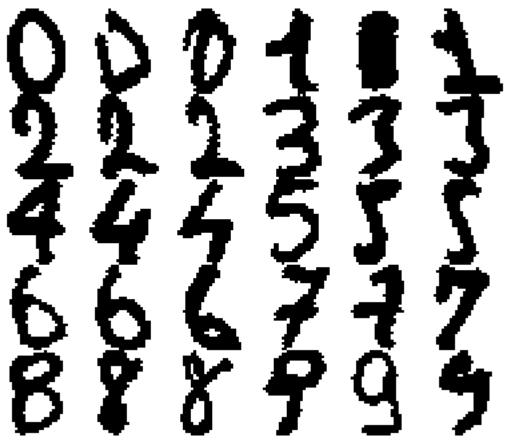

Figure 4: Digits samples.

5. RESULTS AND DISCUSSION

The complete experimental results are in Table 1: for each network run, the column "Im" stores the number of images recognized by the respective class, "Lab." indicates the winner digit label of the class, "Rec.%" is the percentual frequency of the winner digit within that class.

The first run of the Fuzzy ART network is characterized by a single class that, while recognizing over 50% of the image set, its winner digit has a frequency of just above 15%, signaling that this particular class recognised images of other digits (with similar frequencies), and therefore has a poor classification performance; the average winner digit frequencies of this run's classes is also low, pointing to the same situation. The second run of the Fuzzy ART is very similar to the first one, with no significant improvement on the winner digits' frequencies.

The first and second runs of the AUTOWISARD network are characterized by a more even distribution of the images over the existing 20 classes than the Fuzzy ART's. The average winner digit frequencies are also higher, suggesting that the AUTOWISARD's classes do recognize fewer images of different digits. It is also worth noting that, although using the same constant number of classes, in the second AUTOWISARD run all digits were represented at least by one class, contrasting with the second run of the

Fuzzy ART, when the digits 5, 8 and 9 weren't represented individually at all.

The better classifications generated by the AUTOWISARD network are explained by its richer internal class representation, inherited from the original WISARD network. As the Fuzzy ART uses a single vector to represent the class (that is, a single point in the $\{0,1\}^n$ space), the (AUTO) WISARD uses a discriminator instead, which can represent simultaneously a number of these vector (points). The number of different points that can be represented by a discriminator is a function of the diversity of its training patterns, and is a consequence of the monotonic training of the RAM neurons. By example, when a clear m-neuron discriminator is trained with 2 totally different patterns (a solid black and a solid white image), the number of different points stored in it is 2^m.

The capacity of the discriminator to represent multiple points, together with the partial training feature of the AUTOWISARD, enables the creation of more complex class separation surfaces than the single vector approach's, which is reflected in the average radius of the AUTOWISARD's classes that, whilst apparently being larger than the Fuzzy ART's, does not imply in worse classifications (larger clusters).

6. CONCLUSIONS

The AUTOWISARD is a new and powerful extension to the WISARD weightless neural network, which was developed by making judicious use of so-called limitations of the simpler weightless neural networks, as its monotonic training and propension to achieve a state of saturation, to bring unsupervised learning characteristics to that original (and also powerful) model. In a sample application using complex (and large) patterns, when compared against a similarly featured neural network (Fuzzy ART), the AUTOWISARD was able to present better, more realistic classifications, without growing more classes than its counterpart. The AUTOWISARD is a viable and useful alternative to the more mainstream unsupervised neural classifiers, expanding the universe of WISARD applications.

7. ACKNOWLEDGEMENTS

This work was supported by a research grant from CNPq - Brazilian Research Council.

8. REFERENCES

[1] Wickert, I., França, F. M. G.: AUTOWISARD: Unsupervised Modes for the WISARD. *Lecture Notes in Computer Science 2084*, Springer-Verlag, Berlin, pp. 435–441, 2001

[2] Aleksander, I., Thomas, W., Bowden, P.: WISARD, a radical step forward in image recognition. *Sensor Ver.*, pp. 120–124, 1984

[3] Aleksander, I., Morton, H..: *An Introduction to Neural Computing*, Chapman and Hall, London, 1990

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

[5] Grossberg, S.: Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science 11*, pp. 23–63, 1987

[6] Alpaydin, E., Kaynak, C.: Bogazici University, Turkey. ftp://ftp.ics.uci.edu/pub/machine-learning-databases/optdigits/ (1998)

	Fuzzy ART (1)				Fuzzy ART (2)				AUTOWISARD (1)				AUTOWISARD (2)			
Class	Im.	Lab.	Rec.%	Rad.	Im.	Lab.	Rec.%	Rad.	Im.	Lab.	Rec.%	Rad.	Im.	Lab.	Rec.%	Rad.
1	1134	3	15.43	392.00	856	0	18.93	396.14	51	5	41.18	618.92	56	1	69.64	633.55
2	74	2	40.54	400.00	253	4	16.21	398.76	18	1	66.67	690.39	42	8	83.33	606.76
3	59	7	47.46	395.85	63	2	36.51	391.70	87	9	54.02	604.33	28	7	85.71	545.36
4	25	4	36.00	388.32	109	2	24.77	385.60	45	9	71.11	605.60	51	9	80.39	541.29
5	22	2	18.18	410.55	41	4	36.58	398.32	32	1	46.88	665.19	57	8	89.47	657.47
6	76	2	55.26	379.76	36	2	41.67	404.47	129	6	89.15	595.94	15	5	80.00	504.20
7	25	4	44.00	400.56	37	7	37.84	402.76	92	4	90.22	579.41	16	7	100.00	577.44
8	39	6	33.33	394.20	16	2	37.50	394.75	18	2	77.78	610.17	89	1	69.66	607.15
9	18	6	27.77	409.72	30	4	36.67	384.47	35	4	100.00	546.63	84	5	92.86	595.81
10	15	4	40.00	399.53	24	2	33.33	396.29	96	9	58.33	640.16	34	5	73.53	626.68
11	20	4	20.00	382.15	30	4	53.33	393.70	187	0	97.86	551.33	203	6	75.86	580.20
12	27	6	33.33	392.93	16	6	62.50	390.56	30	7	83.33	683.60	76	4	96.05	564.11
13	24	6	25.00	396.46	19	6	73.68	410.68	172	5	80.81	603.42	37	9	37.84	618.16
14	14	5	28.57	394.21	17	7	58.82	399.59	37	4	97.30	572.35	21	8	47.62	683.95
15	12	7	41.66	394.83	15	6	40.00	376.33	102	1	70.59	551.34	133	4	63.16	614.50
16	22	6	27.27	375.09	120	7	35.00	412.18	94	8	77.66	704.39	158	2	84.18	706.04
17	10	4	70.00	373.30	16	4	62.50	367.88	27	9	55.56	683.48	40	9	80.00	535.88
18	181	7	24.31	410.25	22	1	50.00	416.23	216	7	74.54	581.99	302	0	62.25	593.84
19	133	1	48.12	391.74	206	3	37.86	408.48	378	2	46.03	680.89	246	7	63.82	646.32
20	4	1	25.00	294.50	8	6	50.00	375.38	88	6	86.36	583.23	246	3	70.33	609.06

Table 1: The complete experimental results. ”Im.” is the number of images recognised by that class, ”Lab.” is the label assigned to it, ”Rec.%” is the frequence of the images of that label within that class and ”Rad.” is the class’ mean Hamming radius, in bits.