AUTOWISARD: Unsupervised Modes for the WISARD

Iuri Wickert and Felipe M. G. França

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

Abstract. This work introduces two new unsupervised learning algorithms based on the WISARD weightless neural classifier model. The first one, the standard AUTOWISARD model, is able to perform fast one-shot learning of unsorted sets of input patterns. The second one is a recursive version of AUTOWISARD which not only keeps the good features of the basic model, agility, plasticity and stability, but also produces hierarchically structured tree-like classifications. Although the standard AUTOWISARD model exhibits good classification skills when exposed to symbolic patterns (by producing only few classes containing patterns of different meanings), its recursive version produces even less confusing classes. The stability of both learning algorithms is also demonstrated.

1 Introduction

The WISARD is a well-known weightless neural network model, based on a very simple neuron model (RAM) and architecture (class discriminators) [1], [2]. Despite its simplicity, this model provides good generalisation and is able to perform fast, one-shot learning of binary input patterns. The main purpose of this work is to introduce new unsupervised learning algorithms for this neural network model, with the intention to bring together the good features of both WISARD and ART models [3].

WIS-ART, a hybrid approach where both models coexist, was firstly introduced by Fulcher [4]. The new AUTOWISARD models introduced here do not include the ART model explicitly, like in WIS-ART, but have embedded the essence of its behaviour within the WISARD model. Moreover, as it will be shown later in this text, the ability of the ART model on representing classes of patterns through vector quantisation (stylisation) is generalised in the AUTOWISARD models in the sense that each class can be defined by a family of vectors. This may lead to a greater richness in terms of the topology of the frontiers among the different resulting classes.

The following is how the remainder of this paper is organised. A quick overview of the WISARD and ART models is given in the next section. Section 3 introduces the basic AUTOWISARD model, including the proof of its stability, what is true for both AUTOWISARD models. Next, in Section 4, the recursive version of AUTOWISARD is described. The natural generation of tree-like structured classifications over binary input patterns is demonstrated and its performance analysed. The experiments carried out are presented in Section 5. Section 6 includes a discussion over the obtained results. The conclusions are finally presented in Section 7.

2 The WISARD Neural Network

WISARD (Wilkie, Stonham and Aleksander's Recognition Device) [1], [2] is a multi-discriminator weightless neural paradigm originally designed to perform image pattern recognition. Each discriminator is composed by multiple n-input RAM-type memory units (n-tuples or simply *neurons*) having their outputs connected to a summing device. Before any kind of training occurs, all positions of all tuples of all discriminators are set to 0.

The training phase of a discriminator starts by presenting randomly chosen subsets of the pixels of the entire target binary input image to all n-tuples in such a way that the whole input image is covered. The mapping of the connection of all pixel subsets to the n-tuples will be kept for further training and recognition phases. As illustrated by Fig. [1], each subset of pixels of the input image forms an address defining the n-tuple position to be set to 1.

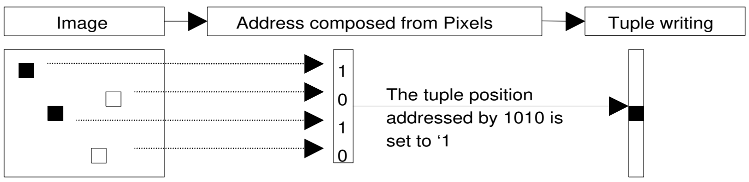

Fig. 1. A Boolean neuron.

The training phase of a discriminator is finished upon the presentation of all different input images that will define its representative class during the recognition phase. After the training of all descriminators, the recognition phase can be started by simultaneously presenting a target input image to all discriminators and comparing their responses r, i.e., the sums of all neurons' outputs belonging to each discriminator. Fig. [2] illustrates the recognition process. Observe that d, the difference between the two best responses, can be used to define a positive class identification. For instance, the relative *confidence* $c = d/r_{\text{best}}$ is usually a good practical choice.

3 The AUTOWISARD Model

The standard AUTOWISARD model is an unsupervised learning algorithm for the WISARD network, which allows a single-pass clustering of unsorted and

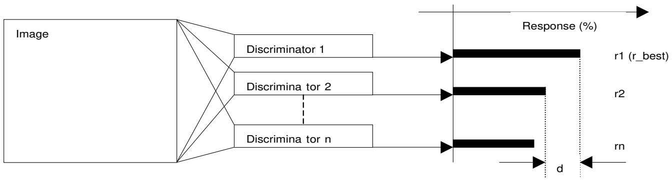

Fig. 2. The WISARD multi-discriminator system.

un-normalized, raw binary data. The main concepts behind AUTOWISARD are the *learning window*, *partial learning* and *semi-probabilistic allocation of classes* (discriminators).

At the beginning, AUTOWISARD can be seen as a WISARD which trains the winner class with a well-matching input sample and allocates dynamically a new class when an input sample couldn't be acceptably recognised by the existing ones. Although this is a workable stategy, one can see that it leads to saturation of the discriminators and doesn't reach stability. Otherwise, if the learning clause above was removed, it can be thought as a simple vector quantization model.

To ensure that AUTOWISARD would reach stability, the allocation of new classes is ruled by a learning window policy. It consist of a recognition interval given by $w_{\min}$ and $w_{\max}$, $0 \leq w_{\min} \leq w_{\max} \leq r_{\max}$, $r_{\max}$ being the number of neurons (RAM nodes) of a discriminator (Fig.[3]).

The learning window policy operates as follows: given r_{best} the WISARD's best recognition value for an input sample, if

- $0 \leq r_{\text{best}} \leq w_{\min}$, a new discriminator is allocated and trained with that sample;
- $w_{\min} < r_{\text{best}} < w_{\max}$, it can either allocate a new class, or submit the winner class to a partial training, probabilistically;
- $w_{\max} \leq r_{\text{best}}$, nothing is done.

The learning window ensures that no well-fitted sample will disturb its class with any extra training, and that ill-fitted samples will get new classes to represent themselves. The cases "in between" (inside the learning window) uses a probability function of the relative position of the recognition inside the window, to choose to allocate a class or to train an existing one. If a r_{best} falls inside the window, but nearer to $w_{\min}$, the probability of allocating a class is greater than to train the existing winner; on the other side, if r_{best} is nearer to $w_{\max}$, then the probability of the winner class be trained is greater. This probabilistic feature was thought to allow a sample that (wrongfully) felt inside the window to have a chance to create a new class to itself and not to contribute to the best class' representation. It was called "semi-probabilistic" because the probabilistic behavior shows only when $w_{\min} < r_{\text{best}} < w_{\max}$. In an attempt to control, or to minimize the effects of saturation on the discriminators, a partial learning method was developed. It consists of training a discriminator just enough to make it able to recognise that sample ($r_{\text{best}} = w_{\max}$). The method estimates the number of

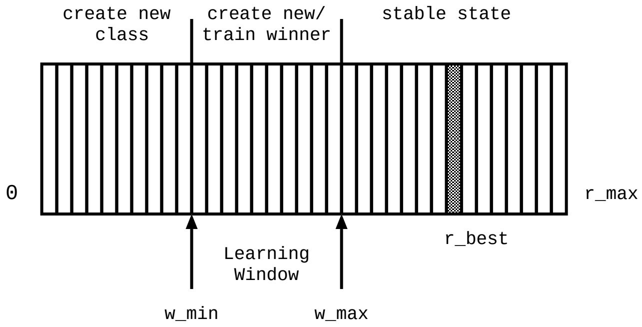

Fig. 3. The Learning Window.

neurons to be taught, $n = w_{\max} - r_{\text{best}}$, randomly selects n neurons from a list of mismatching neurons and trains them according to the input sample.

The learning window policy, associated with the partial training method and the monotonicity of the discriminator's recognition function, leads to a network that stabilizes with an order-less, single-pass training: for a set of training input samples, once the network was trained (in a single pass), it is guaranteed that any of them will have a minimum recognition of $w_{\max}$.

4 The Hierarchical AUTOWISARD

The hierarchical AUTOWISARD is a recursive, tree-like structure using AUTOWISARDs as nodes. Its aim is to recursively find a clustering hierarchy inside the data set, creating sucessive levels of even more specialised (discriminative) AUTOWISARDs, using a *grouping metric* to determine when a sufficiently fine clusterization was achieved. The grouping metric used is the recognition interval of the samples assigned to a discriminator: it is assumed that when a discriminator has a recognition interval above an arbitrary threshold, it is very likely to contain more than one data cluster.

The recursive method to create the hierarchical structure is to link a new AUTOWISARD network to each discriminator that has a large recognition interval, until the recognition intervals of all the discriminators (at any level) are below that threshold or a maximum number of levels is reached. Each network is trained without supervision only with the set of input samples assigned to the parent discriminator. It is desirable to have a mechanism to control the (free) growing of the network, as the creation of deep, super-specialised AUTOWISARDs, tend to present a vector-quantization behavior.

5 Experiments

The example application for the classification abilities of the AUTOWISARD models is the classification of handwritten digits, represented as binary images

($32x32$ pixels), whose symbolic classes are previously known. The digits database employed is the training set of "Optical Recognition of Handwritten Digits" database [5], in its un-normalized version, freely available.

The input image set has the following digits distribution:

Digit	0	1	2	3	4	5	6	7	8	9
Qty.	189	198	195	199	186	187	195	201	180	204

The experiments consists of training a set of standard AUTOWISARD networks with the unsorted images and to compare the clusters created inside each network for *symbolic and quantitative confusion*, and *overfitting*.

Network	Classes	Classes with multiple symbols				Winners' recognition average.(%)	Overfitting (%)
		2	3	4	5+		
1	126	30	7	7	0	80.53	75.40
2	140	29	14	6	1	75.27	80.00
3	131	21	13	4	2	77.49	76.34
4	136	28	11	4	3	79.84	75.00
5	133	35	5	2	2	80.81	76.69
6	135	27	5	5	1	79.98	77.04
7	116	25	6	6	4	78.91	77.59
8	139	30	13	5	0	79.82	74.82
9	133	33	10	6	0	80.26	72.18
10	123	28	7	3	1	80.25	78.05
11	117	27	11	2	1	80.38	73.50
12	133	29	8	7	1	79.28	75.19
13	130	26	15	4	1	80.31	76.15
14	144	28	12	5	1	77.49	79.17
15	134	27	5	5	1	79.01	79.85
16	128	23	14	5	0	81.07	67.97
17	126	33	11	5	0	76.47	73.81
18	122	37	12	4	0	78.78	74.59
19	138	29	7	4	2	76.96	77.54
20	122	30	8	5	3	82.38	71.31
AVG.	130.3	28.75	9.70	4.70	1.20	79.26	75.61
STDEV	7.49	3.70	3.24	1.34	1.12	1.70	2.96

Table 1. The standard AUTOWISARD results.

The symbolic and quantitative confusion are measured as the percentile of classes that contains multiple symbols (digit labels) and, for that classes, the mean percentile of the winner symbol's stances, respectively, and were meant to characterize the quality of (symbolic) classification and to search for classification "trends" inside the ambiguous discriminators. The overfitting measure is the percentile of classes that represents a smaller number of stances than an overfitting threshold, giving feedback of the network's generalisation performance.

The hierarchical AUTOWISARD test consists of a single training phase with a limited maximum number of levels. For matters of space, only a branch of the hierarchy will be shown.

The experiments were realized with fixed set of parameters (tuple size, learning window parameters and probability distribution function), because the analysis of the AUTOWISARD sensibility to this parameters was beyond the scope of this paper. The experiments were thought to bring to the readers a feeling of the models capacities rather than to be an authoritative, extensive models benchmarking

6 Results and Analysis

A set of 20 AUTOWISARDs were trained with the same set of randomly ordered input images, with 6-bits tuples, $w_{\min} = 30\%$, $w_{\max} = 40\%$. The overfitting threshold is 1%: an overfitted class represents 19 or less images, from a set of 1924 images. The results are shown in Table [1]. It also shows the mean and standard deviation of these results, illustrating that, in spite of training the networks with random ordered samples and using different input mappings, the networks shows classification profiles qualitatively similar to each other's.

The low deviation of the winner's average on the classes with more than one symbol indicates that, even when there is confusion inside a class, still exists a clear winner digit and that these discriminators wasn't driven into saturation (when it wouldn't be possible to point a winner digit).

The hierarchical AUTOWISARD was trained with the same set of images used above, with a maximum of 2 layers (using tuples of 4 and 8 bits, respectively), $w_{\min} = 30\%$, $w_{\max} = 50\%$, and a subclassing threshold of 35%. A selected branch is shown in Fig. [4]. Observe that, as the sub-networks were trained only with their parents' pre-classified images, it is easier to classify them into a more specialised way.

7 Conclusions

The new AUTOWISARD unsupervised learning approaches, based on a simple weightless neural model, were introduced and their classification potentialities identified. An interesting feature offered by AUTOWISARD lies in its implicit recognition phase ($w_{\max} < r_{\text{best}}$) during training. A potential advantage of that is the greater spacial stability of the clusters created by AUTOWISARD's constructive approach, thanks to the (recognition) monotonicity provided by the basic WISARD model.

The experimental results have shown a strong consistency on the formation of classes, as indicated by (i) the number of hierarchical clusters generated under different trials, and; (ii) that inside the few classes formed by patterns of different

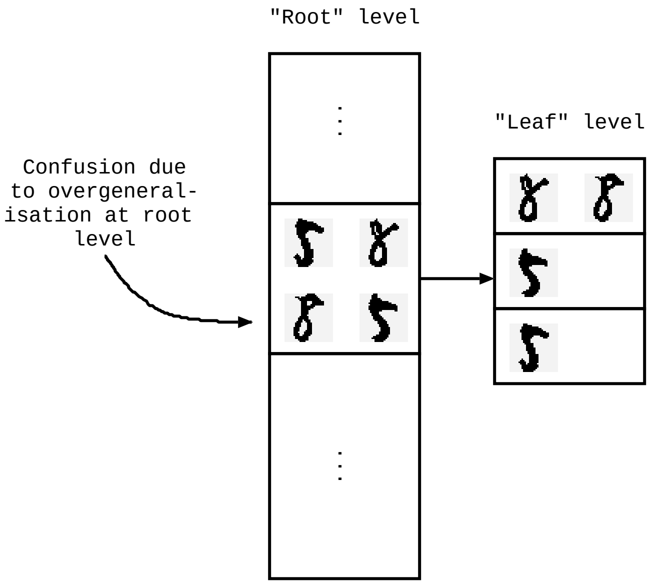

Fig. 4. Better classification through (hierarchical) specialization

symbolic meanings, it is easy to evidentiate a winner, i.e., wrongly classified patterns are a minority inside such classes.

8 Acknowledgements

The authors would like to thank Sergio Exel for the helpful comments, and the grant support from CNPq - Brazilian Research Council.

References

[1] Aleksander, I., Thomas, W., Bowden, P.: WISARD, a radical step forward in image recognition. Sensor Ver., (1984) 120–124
[2] Aleksander, I., Morton, H..: An Introduction to Neural Computing. Chapman and Hall, London (1990)
[3] Grossberg, S.: Competitive learning: From interactive activation to adaptive resonance. Cognitive Science **11** (1987) 23–63
[4] Fulcher, E.: WIS-ART: Unsupervised clustering with RAM discriminators. Intl. J. of Neural Systems, **3**, n. 1 (1992) 57–63
[5] Alpaydin, E., Kaynak, C.: Bogazici University, Turkey. ftp://ftp.ics.uci.edu/pub/machine-learning-databases/optdigits/ (1998)