Neurocomputing 73 (2010) 1057–1064

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Producing pattern examples from "mental" images ☆

Bruno P.A. Grieco [a], Priscila M.V. Lima [b], Massimo De Gregorio [c], Felipe M.G. França [a,b,*]

[a] PESC/COPPE, Universidade Federal do Rio de Janeiro, Caixa Postal 68511, 21941-972 Rio de Janeiro, RJ, Brazil
[b] LAM—Computer Architecture and Microelectronics Laboratory, Universidade Federal do Rio de Janeiro, Caixa Postal 68511, 21941-972 Rio de Janeiro, RJ, Brazil
[c] Istituto di Cibernetica Eduardo Caianiello, CNR, Via Campi Flegrei 34, 80078 Pozzuoli, NA, Italy

ARTICLE INFO

Available online 21 January 2010

Keywords:
WiSARD
DRASiW
Mental images
Pattern recognition
Pattern generation
Weightless neural networks

ABSTRACT

The WiSARD (Wilkie, Stonham and Aleksander's Recognition Device) weightless neural network model has its functionality based on the collective response of RAM-based neurons. WiSARD's learning phase consists on writing at the RAM neurons' positions addressed (typically through a pseudo-random mapping) by binary training patterns. By counting the frequency of writing accesses at RAM neuron positions during the learning phase, it is possible to associate the most accessed addresses with the corresponding input field contents that defined them. The idea of associating this process with the formation of "mental" images is explored in the DRASiW model, a WiSARD extension provided with the ability of producing pattern examples, or prototypes, derived from learnt categories. This work demonstrates the equivalence of two ways of generating such prototypes: (i) via frequency counting and filtering and (ii) via formulating fuzzy rules. Moreover, it is shown, through the exploration of the MNIST database of handwritten digits as benchmark, how the process of mental images formation can improve WiSARD's classification skills.

1. Introduction

Although very important to many applications of intelligent systems, explaining the resulting choice of a category is a process that may prove hard to accomplish. In particular, in intelligent computational models, such as feedforward neural networks, where knowledge is stored in a distributed fashion (i.e., synaptic weights hold the results of the learning/training process) and pattern classification is performed via an input-to-output unidirectional flow, it is hard to obtain explicit information about *which* pattern(s) the classifier internally associates with a target output. In other words: how the classifier could produce a pattern example, a *prototype*, of a particular class? Apart from being a kind of explanation coming from an artificial classifier, another question would be about the usefulness of such information. This paper presents qualitative and quantitative explorations of a weightless neural system capable of producing prototypes, i.e., self-generated pattern examples. Besides, positive results concerning the use of prototypes as a disambiguation tool in the classification process are discussed.

The pioneering use of *n*-tuple RAM nodes in pattern recognition problems is due to the work of Bledsoe and Browning in the late 1950s [18]. A few years later, Aleksander introduced the stored logic adaptive microcircuit (SLAM), i.e., *n*-tuple RAM nodes as basic components for an adaptive learning network [19]. Created by Wilkes, Stonham, and Aleksander in 1984, the WiSARD perceptron was the first artificial neural network machine (and the most representative weightless neural network (WNN) model) to be patented and produced commercially [1]. Many other WNN paradigms were proposed and have been surveyed in [20,13]. WiSARD takes a set of bits as input, which is then parsed into a set of uncorrelated *n*-tuples. Each *n*-tuple is used as a specific address of a RAM-based neuron, in such a way that the input field is completely covered. A WiSARD *discriminator* is composed by a set of *n*-tuples covering the whole input field, and is trained with representative data of a specific class/category. A discriminator recognizes a test pattern via summation of all of its associated RAM neurons' output. Therefore, the WiSARD model is a multi-discriminator, unidirectional architecture, in summary, a perceptron.

The DRASiW model introduces a way of providing the WiSARD model with backward-classification capabilities, so that one can ask for prototypes of already learnt categories [4,7], i.e., each discriminator is able to produce representative examples of a class that have been learnt from trained patterns. In order to make this possible, RAM neurons' positions act as access counters, which contents can be reversed to an internal "retina", where a "mental" image is produced, thus yielding a bidirectional structure. The "mental" image metaphor, associated with the

☆ Funded by: BZG.
* Corresponding author at: PESC/COPPE, Universidade Federal do Rio de Janeiro, Caixa Postal 68511, 21941-972 Rio de Janeiro, RJ, Brazil. Tel.: +55 21 2562 8663; fax: +55 21 2562 8676.
E-mail addresses: bgrieco@mac.com (B.P. Grieco), priscilamvl@gmail.com (P.M. Lima), m.degregorio@cib.na.cnr.it (M. De Gregorio), felipe@cos.ufrj.br (F.M. França).

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2009.11.015

internal "retina" metaphor, was originally explored in [17], which also discusses the cognitive plausibilities related to these ideas. Prototypes, or pattern examples, of trained classes, are obtained from a mental image by simply selecting RAM neurons' positions with highest values.

Fuzzy logic has improved weightless neural network performance on categorization and cognition [8,9] and both Boolean [10] and fuzzy rule [11] extraction has been explored in this context. Based on the preliminary ideas presented in [14], this work proposes an equivalent way of expliciting learnt categories, based on the same information considered in the DRASiW model, in the form of fuzzy rules. The MNIST database of handwritten digits [15] was used to demonstrate how both DRASiW and the equivalent fuzzy rule formulation are able to produce quite representative pattern examples. Moreover, as an interesting byproduct, it is also shown how the process of mental images formation provided by DRASiW can improve WiSARD's classification skills.

The next section explains the WiSARD weightless neural network model. Section 3 describes how "mental" images are produced by the DRASiW generalization. Section 4 demonstrates an equivalent fuzzy rule formulation. The generation of pattern examples by both DRASiW and its equivalent fuzzy rule specification is explored and compared, via experimentation, in Section 5. A novel way to improve WiSARD's classification capabilities is introduced and tested in Section 6. Section 7 presents our conclusions.

2. The WiSARD perceptron

A RAM-discriminator consists of a set of X one-bit word RAM nodes, or RAM neurons, with n inputs and a summing device (Σ) [1]. Any such RAM-discriminator can receive a binary pattern of ($X \times n$) bits as input. The RAM input lines are connected to the input pattern by means of a biunivocal pseudo-random mapping (see left part of Fig. 1). The summing device enables this network of RAM nodes to exhibit—just like other ANN models based on synaptic weights—generalization and noise tolerance [2].

In order to train the discriminator, one has to set all RAM memory locations to "0" and choose a training set formed by binary patterns of ($X \times n$) bits. For each training pattern, a "1" is stored in the memory location of each RAM addressed by this input pattern. Once the training of patterns is completed, RAM memory contents will be set to a certain number of "0"s and "1"s.

The information stored by RAM nodes during the training phase is used to deal with previous unseen patterns. When one of these is given as input, RAM memory contents addressed by the input pattern are read and summed by Σ. The number r thus obtained, which is called the *discriminator response*, is equal to the number of RAMs that output "1". It is easy to see that r necessarily reaches the maximum X if the input pattern belongs to the training set. r is equal to "0" if no n-bit component of the input pattern appears in the training set (not a single RAM outputs "1"). Intermediate values of r express a kind of "similarity measure" of the input pattern with respect to the patterns in the training set.

A system formed by various RAM-discriminators is called WiSARD (Wilkie, Stonham and Aleksander's Recognition Device) [1,13]. Each RAM-discriminator is trained upon a particular class of patterns, and classification by the multi-discriminator system is performed in the following way. When a pattern is given as input, each RAM-discriminator gives a response to that input. The various responses are evaluated by an algorithm which compares them and computes the relative confidence c of the highest response (e.g., the difference d between the highest response and the second highest response, divided by the highest response). A schematic representation of a RAM-discriminator and a 10 RAM-discriminator WiSARD are illustrated by Fig. 1.

The performance of the WiSARD strongly depends on n. Other factors, such as the choice of the training set, the way confidence is calculated, etc., also influences WiSARD's performance. Specialized responses from WiSARD grows with n; on the other hand, generalization capabilities of WiSARD grows inversely with n [2].

3. DRASiW: WiSARD's "mental" images

DRASiW is an extension to the WiSARD model provided with the ability of producing pattern examples, or prototypes, derived from learned categories. RAM-discriminators are modified in what their memory locations may hold and, correspondingly, in their training algorithm. These changes, which produce something very similar to PLN nodes, introduced by Aleksander [3], allows one to store q-bit words in memory locations (where q is usually not greater than 8); this information, in turn, can be exploited in the generation of "mental" images of learned pattern categories (further improving in other ways the behaviour of RAM-discriminators).

The training algorithm of RAM-discriminators is changed in one aspect only: instead of storing "1"s, it just increment (+1) memory location contents that are addressed by input patterns. At the end of the training phase, values of the memory contents will vary between 0 and Y (where Y is the number of training patterns). Fig. 2 shows the result of training the same RAM-discriminator of Fig. 1 with examples of stylized vertical lines (Fig. 2 left), by means of the new algorithm.

The various memory content values can now be associated with subpattern frequency in the training set. For instance, the memory content of the address 010 associated with the $\diamond$−RAM is 5. This value indicates that the subpattern 010 is present 5 times in the training set of Fig. 2. One should notice that the new domain of memory content values (non-negative integers) do not

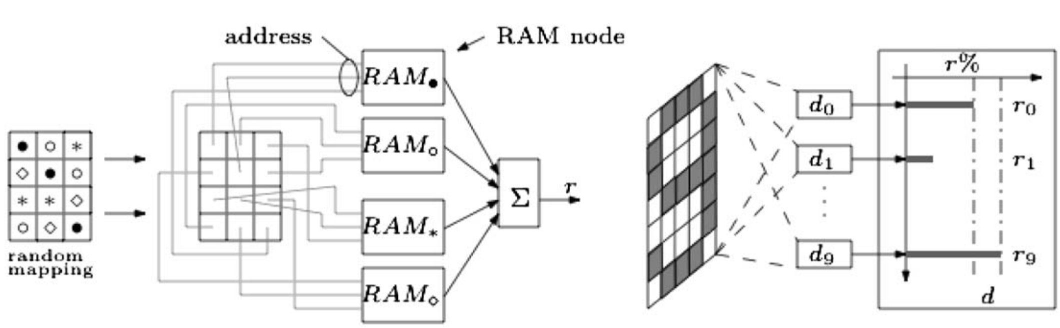

Fig. 1. Example of a RAM-discriminator (left) and of a 10 RAM discriminator WiSARD (right).