Can the false-discovery rate be misleading?

Rodrigo Barboza[1], Daniel Cociorva[2], Tao Xu[2], Valmir C Barbosa[1], Jonas Perales[4], Richard H Valente[4], Felipe M G França[1], John R Yates III[2], Paulo C Carvalho[3,4]

1. Systems Engineering and Computer Science Program, COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
2. Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California, USA
3. Center for Technological Development in Health of the Oswaldo Cruz Foundation, Rio de Janeiro, Brazil
4. Laboratory of Toxinology, Oswaldo Cruz Institute, Rio de Janeiro, Brazil

Supplementary material

Overview of the discriminant methods

Briefly, both strategies employed in this work rely on a number of discriminating functions equal to that of the classes for the problem at hand. In our case, a dichotomy, the strategies use two discriminant functions,

$$g_i(\vec{x}), i \in \{target, reverse\}.$$

The confidence score *DS* is computed as the difference in scores from the discriminators of the forward and reverse classes:

$$DS = g_{target}(\vec{x}) - g_{reverse}(\vec{x}).$$

A – The Bayesian discriminant function

Each Bayesian discriminant function is

$$g_i(\vec{x}) = -\frac{1}{2}(\vec{x}-\vec{\mu}_i)^t \Sigma_i^{-1}(\vec{x}-\vec{\mu}_i) - \frac{1}{2}\ln|\Sigma_i| + \ln(P(\omega_i)),$$

where the prior probability for a class, $P(\omega)$, is assumed to be 0.5. In the above, $\vec{\mu}$ is the mean vector of the training matrix, Σ is the covariance matrix, $|\Sigma|$ its determinant, and Σ^{-1} its inverse.

B - The WiSARD weightless artificial neural network model

Mainstream artificial neural network models are inspired on synaptic strengths, implemented as artificial weighted-sum-and-threshold neurons, e.g., McCullogh and Pitts' paradigmatic neuron stylization[1]. Weightless neural networks have, as biological inspiration, the way a neuron's dendritic tree (DT), a quite noticeable morphological structure of the neuron cell, processes excitatory and inhibitory signaling coming from pre-synaptic neurons[2]. A simple analogy does exist between the signaling priority associated with the height of the connections of pre-synaptic neurons to the DT (synapses closer to the neuron's soma have higher priority since they are able to modulate the influence of farer connections) and how address binary decoding is performed in Random Access Memories (RAMs).

The WiSARD (**Wi**lkes, **S**tonham and **A**leksander **R**ecognition **D**evice)[3] is a weightless neural network based on class discriminators built of RAM-based neurons, or RAM nodes. Each discriminator d_i, consists of *X* one-bit word RAM nodes having *n* address input bits. The input vector $\vec{x}$ has all of its $X \times n$ bits connected, via a one-to-one pseudo-random mapping, to RAM node address inputs. The input vector $\vec{x}$, in a more elaborated form, can be loaded from a set of *s* decimal numbers, e.g., scores, which are transformed into a string of $X \times n$ bits via Gray encoding[4].

In order to train d_i, an input vector $\vec{x}$ containing a pattern example of a target class is presented, via the transformation cited above, to all RAM nodes of d_i. Each memory position addressed by its corresponding input bits is marked as "visited", i.e., a '1' is written (all RAMs have all their contents set to '0' as initial condition). This process is repeated for each new pattern example that d_i is meant to recognize. Other discriminators, each one representing a different class, are trained in the same way as d_i. A discriminator response $g_i(\vec{x})$ is obtained via the summation of all RAM node outputs belonging to d_i upon the presentation, to all discriminators, of a given test pattern, loaded into $\vec{x}$:

$$g_i(\vec{x}) = \sum_{j=0}^{X-1} R_i^j ,$$

where R_i^j represents the output of the j^{th} RAM node output of discriminator d_i. Figure 1 illustrates a WiSARD architecture oriented to the recognition of alphabetical characters; notice that a confidence score *DS* is defined as the difference between the two highest discriminator responses.

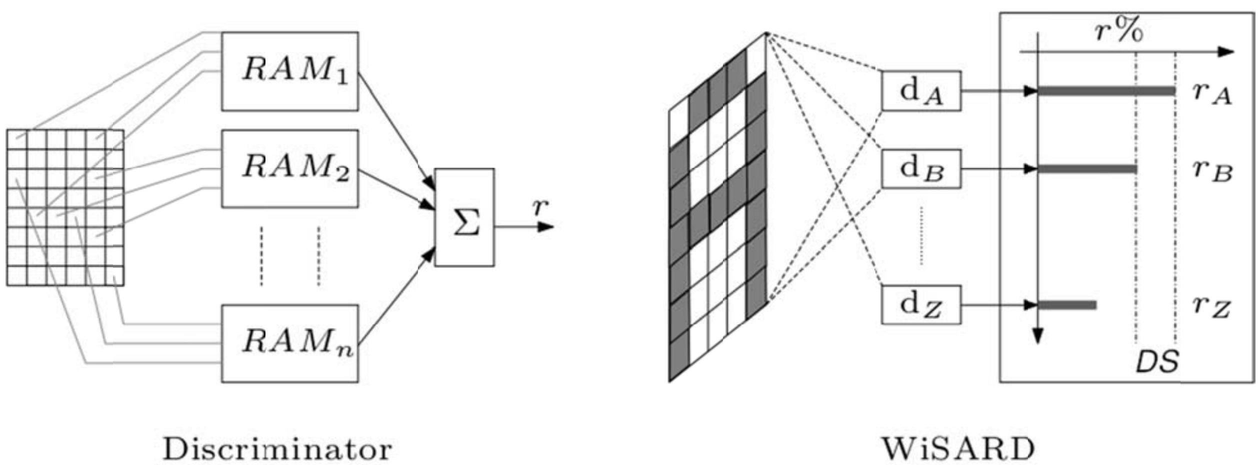

Figure 1. Example of a RAM-based discriminator (left) and a WiSARD having A-to-Z discriminators (right).

As the size of the training set increases, it is expected that two or more discriminators may output the highest value possible (saturation), i.e., $g_i(\vec{x}) = X$, when exposed to test patterns. In order to tackle this problem, a simple generalization of the WiSARD architecture is used[5]: each RAM node position holds a *c*-bit counter so that, at the start of the training process each counter is initialized with value '0' and incremented each time the same position is visited. After the training process is done, counters hold values equal or greater than '0'.

A generalized test pattern procedure uses an integer variable threshold *b*, which is applied over all RAM positions before discriminator responses $g_i(\vec{x})$ are computed. Starting from *b* = 1, discriminator responses are computed; if *DS* is zero, i.e., a draw occurs, *b* is incremented and all discriminator responses are computed again. This procedure is repeated until *DS* becomes greater than zero.

Reference List

1. McCulloch, W. and Pitts, W. *Bulletin of Mathematical Biophysics* 7, 115-133 (1943).

2. Aleksander, I., De Gregorio, M., França, F.M.G., Lima, P.M.V. and H. Morton. *Proc. of the 17th European Symposium on Artificial Neural Networks* (ESANN 2009), Bruges, Belgium, 2009, 299–305.

3. Aleksander,I., Thomas,W., & Bowden,P. *Sensor Rev* 4, 120-124 (1984).

4. Gray, F.: “Pulse Code Communication”; US patent #2,632,058. March 17th, 1953.

5. Grieco,B.P.A., Lima,P.M.V., De Gregorio,M., & França,F.M.G. *Neurocomputing* 73, 1057-1064 (2010).