ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Improving the DRASiW performance by exploiting its own "Mental Images"

Gianluca Coda, Massimo De Gregorio, Antonio Sorgente and Paolo Vanacore

Istituto di Scienze Applicate e Sistemi Intelligenti – CNR

Abstract. Several improvements have been proposed in the literature for the Weightless Neural Networks (WNNs), in particular the DRASiW extension of the WiSARD model with the introduction of mental imagery and bleaching procedure. We propose a new bleaching procedure called Dynamic Adaptive Bleaching (DAB) and its variant, *refined* Dynamic Adaptive Bleaching (rDAB), to improve the WNNs performance in terms of computational time and classification capabilities.

1 Introduction

In the last few years, WNNs have been applied as classifiers in various application domains such as NLP [1, 2], medicine [3, 4], real time video analysis [5] and every time new extensions and/or algorithms [6, 7] have been proposed to improve computational time and classification power. This is in line with today's demand for classifier models that are easy and fast during both training and prediction phases, and produce increasingly accurate results.

For the WNNs, among these improvements of the model, the first most significant was the introduction of the DRASiW extension [8, 9], which introduced the bleaching procedure, i.e. a procedure to deal with tie situations in the classification process by ignoring the sub-patterns the network considers too unusual (that is, very few instances seen during the training phase). Two different methods have already been proposed in the past to improve the bleaching process: confidence search [3] and b-bleaching (binary search) [6] both based on the search of the best bleaching threshold.

In this work we propose a new bleaching procedure (DAB - Dynamic Adaptive Bleaching) and a variant of it (rDAB - *refined* Dynamic Adaptive Bleaching) that significantly improve the performance of the model both in terms of computational time and classification power.

2 DAB – Dynamic Adaptive Bleaching

Bleaching is the process that was introduced in [8] to resolve tie situations, i.e. when two or more discriminators give scores very close to each other. Let D_1 and D_2 be two discriminators and r_1 and r_2, with $r_1 > r_2$, respectively their responses. Let $c_{12} = (r_1 - r_2)/r_1$ be the confidence and τ_c the tolerance (fixed in advance). If $c_{12} < \tau_c$ then we have a tie and the belonging class of the given input cannot be established. In this case to overcome ties the bleaching mechanism is invoked.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Bleaching uses the DRASiW stored information to solve discriminator response ties in the following way: (i) a bleaching threshold variable τ_b is set; (ii) all memory locations with access count greater than or equal to τ_b are set to "1"; remaining ones are set to "0". Starting with $\tau_b = 0$, the threshold is increased while there are ties in the discriminators' responses. When there are no more ties, that is $c_{12} > \tau_c$, the class chosen is the one whose associated discriminator has the highest output.

The effect of bleaching on overtraining consists of making the RAM-neurons ignore the sub-patterns that the network considered too uncommon, i.e., the ones that were presented to it less than τ_b times. The bleaching procedure can be very time consuming. In fact, the time to solve ties increases with the dimension of the dataset instances causing DRASiW performance degradation.

In order to optimise the bleaching mechanism performance, a new bleaching strategy has been introduced.

2.1 DAB algorithm

A slight modification of the training algorithm has been made in order to take advantage of the new proposed bleaching procedure we are going to introduce. The standard bleaching procedure is based on the memory location contents. Every DRASiW input is splitted in sub-patterns and each of them forms a binary input to address a RAM-neuron memory location whose content is incremented by "1". How many times a certain sub-pattern has been given in input to the system is the meaning associated to the values of the memory contents. The sub-pattern associated to an empty memory location has never been given in input to the system. With this information, the system can generate a pictorial representation (a gray level image called "Mental Image" – MI) of the information acquired during the training phase. The MI represents the discriminator class "prototype". The darker the pixel the more important is the pixel for its belonging class.

In order to introduce the new proposed bleaching procedure (algorithm 1), let consider a DRASiW formed by two discriminators, D_1 and D_2, and let the confidence c_{12} be smaller than τ_c.

The discriminator response r is the number of non empty memory locations

Algorithm 1 DAB algorithm

1: $\tau_c \leftarrow 0.01$
2: $r_{L_1} = |L_1|/R; \quad r_{L_2} = |L_2|/R$
3: **if** $r_{L_1} > r_{L_2}$ **then**
4: $\quad c_{12} = (r_{L_1} - r_{L_2})/r_{L_1}$
5: **else**
6: $\quad c_{12} = (r_{L_2} - r_{L_1})/r_{L_2}$
7: **end if**
8: **if** $c_{12} \geq \tau_c$ **then**
9: $\quad$ **if** $r_{L_1} > r_{L_2}$ **then**
10: $\quad\quad$ **return** class C_1
11: $\quad$ **else**
12: $\quad\quad$ **return** class C_2
13: $\quad$ **end if**
14: **else**
15: $\quad \tau_b = max(e_{11}, e_{21})$
16: $\quad$ remove $e_{ij} < \tau_b$ from L_i
17: $\quad$ repeat step 2
18: **end if**

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

selected by the given input divided by the number R of RAMs. This is because, for each RAM, only one memory location will be addressed by the input. The contents of all non empty R memory locations are stored and sorted in ascending order in the vectors L_j, where j refers to the respective discriminator D_j. So doing, the first element of each vector (let say e_{21} for the discriminator D_2) represents the memory location with the lowest value. τ_b is set to the highest between the e_{j1}. Then, for each list L_j all values less than or equal to τ_b are removed. The response of each discriminator D_j is given by $r_{L_j} = |L_j|/R$ where $|L_j|$ is the cardinality of L_j. The system keeps recalculating the responses with the new τ_b until the confidence is greater than τ_c.

With this procedure, that can be easily extended to any number of discriminators, the system always choose the best τ_b, drastically reducing the computational time.

2.2 *r*DAB – *r*efined Dynamic Adaptive Bleaching

The DAB and the standard bleaching take into account only those memory locations greater than zero without taking advantage of the memory location values. In fact, the values stored in the memory locations represent how many time those input sub-patterns have been seen by the discriminator during the training phase. The greater the value the more important is the corresponding sub-pattern. In terms of images, the greater the memory location value the darker are the pixels associated to it. In fact, in the MI generated by the discriminator, one can notice that the most relevant and significant parts of the class prototype are the darkest. This information can help the system improve its classification power. In the following (algorithm 2), we introduce a slight modification to, or better, a refinement of the DAB procedure (called rDAB) exploiting the information content of the MI.

Algorithm 2 rDAB algorithm

$\tau_c \leftarrow 0.01$
$r_{L_1} = |L_1|/R; \quad r_{L_2} = |L_2|/R$
if $r_{L_1} > r_{L_2}$ **then**
  $c_{12} = (r_{L_1} - r_{L_2})/r_{L_1}$
else
  $c_{12} = (r_{L_2} - r_{L_1})/r_{L_2}$
end if
$r_{f_1} = sum(L_1)/|C_1|$
$r_{f_2} = sum(L_2)/|C_2|$
if $r_{f_1} > r_{f_2}$ **then**
  $c_{f_{12}} = (r_{f_1} - r_{f_2})/r_{f_1}$
else
  $c_{f_{12}} = (r_{f_2} - r_{f_1})/r_{f_2}$
end if
if $c_{12} \geq \tau_c$ **then**
  if $r_{L_1} > r_{L_2}$ **then**
    return class C_1
  else
    return class C_2
  end if
else
  if $c_{f_{12}} \geq \tau_{c_f}$ **then**
    if $r_{f_1} > r_{f_2}$ **then**
      return class C_1
    else
      return class C_2
    end if
  else
    $\tau_b = max(e_{11}, e_{21})$
    remove $e_{ji} < \tau_b$ from L_i
    repeat step 2
  end if
end if

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Dataset	Train	Test	Classes	Features	Tics	Retina
Arcene	100	100	2	10000	1024	10240000
Statlog shuttle	43500	14500	7	9	4096	36864
Abalone	3133	1044	3	8	4096	32768
Image seg	210	2100	7	19	1024	19456
Optdigit	3823	1797	10	64	16	1024
Monks 3	122	432	2	6	4	24
LSVT	126		2	310	4096	1269760
Breast tissue	106		6	10	4096	40960
Audit data	776		2	23	1024	23552
Glass	214		6	9	2048	18432
Column 1	310		2	6	1024	6144
Bupa	345		2	6	512	3072

Table 1: Datasets involved in the experiments: *Dataset* names; number of *Train* and *Test* set instances; number of *Classes*; number of *Features*; thermometer resolution (*Tics*); DRASiW input dimension (*Retina = Features*Tics*).

Let r_{f_j} be a further response associated to discriminator D_j. r_{f_j} will take into account the information about sub-pattern "frequency" and it is defined as $r_{f_j} = \frac{1}{|C_j|}\sum_{i=1}^{|L_j|} e_{ij}$, where $|C_j|$ is the number of instances of the training set of the class C_j. Furthermore, $c_{f_{12}} = (r_{f_1} - r_{f_2})/r_{f_1}$ if $r_{f_1} > r_{f_2}$ will be the confidence associated to the new responses r_{f_j} and τ_{c_f} will be the relative tolerance. The system keeps recalculating the responses with the new τ_b until either $c_{12} \geq \tau_c$ or $c_{f_{12}} \geq \tau_{c_f}$.

While the system through the DAB procedure controls how many known sub-patterns are present in the given input, with rDAB the system evaluates even how much important they are.

3 Experiments

The experiments were set up with the aim of evaluating how much the new bleaching strategies (DAB and rDAB) make the DRASiW system more performing in terms of computational time and classification power. 12 datasets of the UCI Machine Learning Repository[1] [10] have been chosen to carry out several experiments. Only for 6 datasets train and test sets are available (see Table 1 for details). Different system configurations have been adopted to run the experiments. In particular, we considered both linear and random mapping and 4, 8, 16, 32 and 64 bits to address the RAM neurons. Furthermore, *ad hoc* thermometer tics have been chosen to better represent the numerical features values. For the datasets without the test sets, the experiments have been carried out with a ten-fold stratified cross-validation.

Four different parameters have been chosen to evaluate the new system performance. Two of them, *B-steps* and *Max B* are strictly related to the evaluation of the new bleaching strategies, while the other two, the *F1-score* and the *Iter./sec*, to evaluate the improvement of the system performance if any. In

[1] https://archive-beta.ics.uci.edu/

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

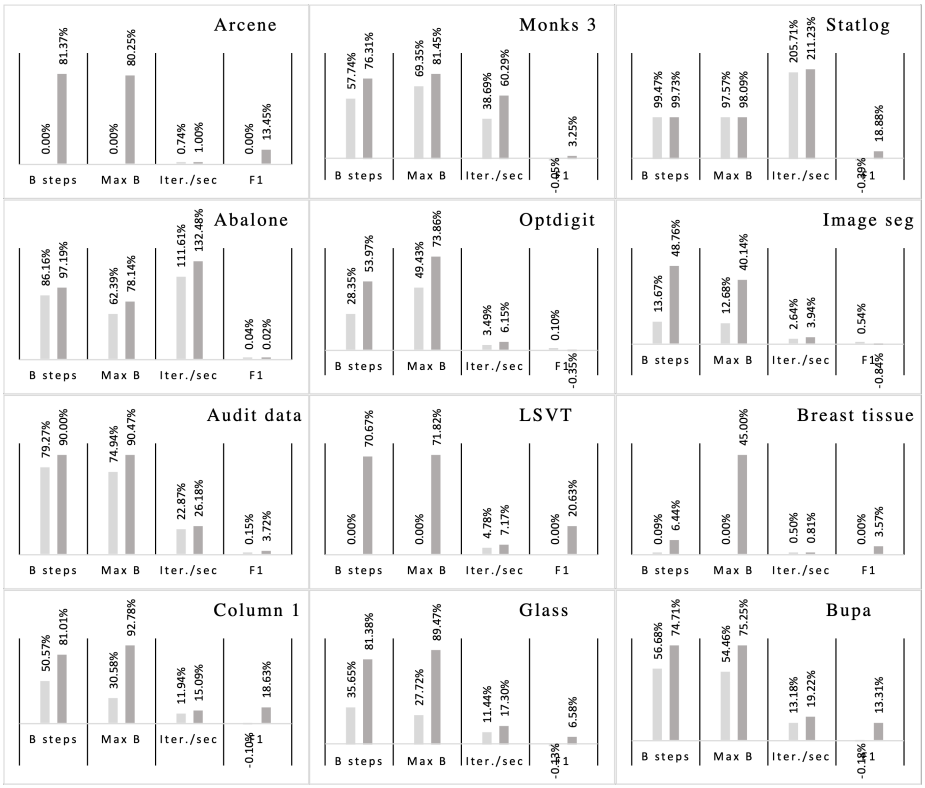

Fig. 1: Percentage of gain achieved by DAB (light gray bar) and rDAB(dark gray bar) with respect to standard bleaching for each dataset

particular and by averaging over all the experiments, *B-steps* is the number of all ties, *Max B* is the maximum threshold values and *Iter./sec* the number of iteration per second.

4 Results

In Figure 1 the results obtaining by applying the new bleaching strategies (DAB and rDAB) are reported. Each single graphic shows the percentage gain achieved by DAB (light gray bar) and rDAB (dark gray bar) with respect to the standard bleaching strategy. The results are very interesting and, just referring to the bleaching parameters, one can notice very high percentage of gain achieved by the new strategies in most of the datasets with different peaks over 90%.

Furthermore, rDAB always improved the system performance (*F1-score*) up the 20% for the *LSVT* dataset. Only for *Optdigit* and *Image seg* a slight drop in performance has been reported (less than the 0.85% on the *F1-score*) but still with a gain in the *Iter./sec*: faster but slightly less performing systems.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

To sum up, average parameter values on all datasets are reported in Table 2.

Parameters	DAB on B	rDAB on B	rDAB on DAB
B steps	42.305%	71.796%	52.740%
Max B	39.927%	76.394%	55.183%
Iter./sec	35.633%	41.738%	4.168%
F1-score	**-0.002%**	8.403%	8.416%

Table 2: Average results on 12 datasets

5 Conclusion

In this work we have presented two new bleaching algorithms and compared them with the standard bleaching by evaluating both the performance accuracy by *F1-score* and the computational time by the number of iterations per second and the number of bleaching steps (*B steps*).

The experiments were performed on twelve datasets and the results show significant improvements on the computational time (see table 2) showing an average gain of 71.796% on the number of ties (*B steps*) and of 41.738% on the number of iterations per second (*Iter./sec*). The tests also showed an improvement in the classification power reaching an average improvement of 8.403%.

References

[1] M. De Gregorio, A. Sorgente, and G. Vettigli. Weightless neural networks for text classification using *tf-idf*. In *ESANN*, pages 239–244, 2021.

[2] Carneiro H.C.C., França F.M.G., and Lima P.M.V. WANN-Tagger-a weightless artificial neural network tagger for the portuguese language. In *IJCCI*, pages 330–335, 2010.

[3] Souza C., Nobre F., Lima P.M.V., Robson S., Brindeiro R., and França F.M.G. Recognition of HIV-1 subtypes and antiretroviral drug resistance using weightless neural networks. In *ESANN*, pages 429–434, 2012.

[4] M. De Gregorio, A. Di Costanzo, A. Motta, D. Paris, and A. Sorgente. Classification of preclinical markers in alzheimer's disease via WiSARD classifier. In *ESANN*, pages 43–48, 2022.

[5] M. De Gregorio and M. Giordano. Background estimation by weightless neural networks. *Pattern Recognition Letters*, 96:55–65, 2017. Scene Background Modeling and Initialization.

[6] Carvalho D.S., Carneiro H.C.C., França F.M.G., and Lima P.M.V. B-bleaching: Agile overtraining avoidance in the WiSARD weightless neural classifier. In *ESANN*, 2013.

[7] Alan T., Bacellar L., Susskind Z., Villon L.A.Q., Miranda I.D. S., de Araújo L.S., Dutra D.L.C., Breternitz Jr. M., LIizy J., Lima P.M.V., and França F.M.G. Distributive thermometer: A new unary encoding for weightless neural networks. In *ESANN*, 2022.

[8] Massimo De Gregorio. On the reversibility of multi-discriminator systems, Technical Report 125/97, Istituto di Cibernetica-CNR, 1997.

[9] C.M. Soares, C.L.F. da Silva, M. De Gregorio, and F.M.G. França. Uma implementação em software do classificador WiSARD. In *V Simpósio Brasileiro de Redes Neurais*, volume 2, pages 225–229, 1998.

[10] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.