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/.

WiSARD-based Ensemble Learning

Leopoldo Lusquino Filho[1], Felipe M.G. França[2] and Priscila M.V. Lima[3],[4] *

1 - São Paulo State University, Institute of Science and Technology, Sorocaba
Av Três de Março, 511 - Alto da Boa Vista, Sorocaba - São Paulo - Brazil
2 - Instituto de Telecomunicações – Porto
Rua Dr. Roberto Frias, w/n, Porto - Portugal
3 - PESC/COPPE 4 - NCE
Federal University of Rio de Janeiro
Av Horácio Macedo, 2030 - Cidade Universitária, Rio de Janeiro - RJ - Brazil

Abstract. Weightless neural networks are recognized for their online learning capacity and competitive performance with the state-of-the-art in different scenarios. Despite this, the literature has not adequately explored the potential of classification ensembles based on these models and their unique characteristics. This study introduces three types of ensembles based on the WiSARD weightless model and evaluates their effectiveness. The results show that these ensembles significantly improve accuracy compared to the WiSARD model and its ClusWiSARD extension, with a reasonable increase in computational cost. Furthermore, using ensembles eliminates the need for time-consuming tie-break policies of traditional WiSARD models.

1 Introduction

Ensemble learning is a very valuable technique in machine learning due to its ability to combine several models in a single committee, in such a way that it tends to have greater accuracy in classification and regression tasks than each of its models individually, even that all of these models have a higher error rate in the ensemble learning. This occurs because the individual models learn fewer samples when training with only one subset[1].

One of the obvious disadvantages of using ensembles is the increase in training and classification time, which makes the use of weightless nets especially recommended for these committees. Despite their great vocation for online learning, WiSARD weightless neural network have been little explored in ensemble learning, and when they have been used, the structure of the committees is strongly oriented to the domain of the problem[2, 3]. Unlike [2, 3], the ensemble types presented in this article can be used for any multi-class supervised learning task in a fully domain-agnostic way. This proposal of new strategies for WiSARD ensembles is corroborated by their exploration in several classification datasets.

*Acknowledgement: This work was partially supported by CAPES, CNPq, FAPERJ and FINEP, Brazilian research agencies, and FCT/COMPETE/FEDER, FCT/CMU IT Project FLOYD: POCI-01-0247-FEDER-045912.

When this research was conducted, the authors Leopoldo Lusquino Filho and Felipe M.G. França were at COPPE/Federal University of Rio de Janeiro.

2 WiSARD neural network

Weightless artificial neural networks whose neurons can be seen as a set of RAM nodes addressed by Boolean inputs and producing Boolean outputs. Among the models that integrate the weightless paradigm, the neural network WiSARD[7] stands out for its simplicity and various extensions. WiSARD is a n-tuple classifier composed of class discriminators. In this model, each discriminator is a set of N RAM nodes, with n addressing lines each. All discriminators share a structure called input retina, which performs a pseudo-random mapping of $N * 2n$ bits from a binary data word to the address lines of all memory locations of the RAM nodes. In the extended version of WiSARD with tie-breaking policies, the content of each RAM address consists of a counter, which is incremented in case of access during the training phase.

As the training of this model consists only of writing in memory, the process is fast and straightforward, enabling online learning. In the classification phase, all discriminators have their RAMs accessed in the respective memory locations addressed by the input bits of the binarized sample, and they all return a score formed by the amount of RAMs accessed in non-null positions. The discriminator with the highest score determines the input class. In case of a tie, a threshold called bleaching is incremented and only memory positions whose counters have values greater than bleaching are computed in the score calculation. WiSARD has an extension called ClusWiSARD[8], which allows the creation of multiple discriminators per class, ensuring that heterogeneity does not compromise its learning. ClusWiSARD employs the traditional WiSARD classification process, differing only in having multiple discriminators for a single class, so ensembling strategies can be applied to it, treating each discriminator as a weak learner.

3 Proposed WiSARD Ensembles

All the ensembles proposed[1]and listed below can be created from a combination of WiSARD, ClusWiSARDs, or both models, which will be called a heterogeneous ensemble. The address size of each neural network is determined randomly. Bagging[4] and Boosting[5] based ensembles were chosen because the efficiency of such methods is proved in the literature. Since the goal was to maintain online training without sacrificing it when using the ensemble. Therefore, the validation stage of the traditional boosting algorithm is not used to create new weak learners, but to assign weights to them according to their performance. WiSARD Borda Count was created in order to avoid the use of the bleaching algorithm on learners, in order to speed up the classification of individual models and to obtain the maximum likelihood estimator of the class ranking in a computationally cheap way.

[1] WiSARD Bagging and WiSARD Boosting proposed here are inspired by the regression ensembles presented in the paper [14] by the same authors of this work and the WiSARD Borda Count is an unpublished strategy originally presented in the PhD thesis of the first author of this work[13].

3.1 WiSARD Bagging

In this ensemble, a set of weightless networks are used as weak learners and learn a subset of the training set and the response of the ensemble consists of the class with the most votes. In the event of a tie, one of the most voted classes is chosen arbitrarily in response to the ensemble. All weak learners are trained using exactly the same number of samples, which can be the same size as the original training set or just a partition of it, which are drawn with replacement. It is also possible to carry out the draw with repetition in the same subset so that the same sample can be trained more than once by the same network.

3.2 WiSARD Boosting

Ensemble composed by weak learners that are trained with subsets obtained from the original training set without replacement and without repetition. The training/validation split is 70%/30% and the vote weight of each model is calculated based on a normalization of its accuracy. Here, the choice of training subsets is not based on validation, being it used only for calculating the weight of the vote of each weak learner in the ensemble's output. The purpose of this ensemble is to increase the degree of differentiation between the models, making them specialists. Its classification stage is exactly like that of WiSARD Bagging.

3.3 WiSARD Borda Count

This ensemble follows the same training process as WiSARD Bagging, but with a different classification system based on diversified voting policies using the Borda count election system[6]. In this ensemble, each learner casts a vote for all class options in the domain based on the scores obtained in its classification phase. The score is determined by the number of non-null memory locations accessed by each of its discriminators. These ranks are called "ballots" and ties are broken arbitrarily. These ballots will be used to calculate the ensemble's response according to any of the following policies, whose classification systems are showed in Table 1: **(i) Starting at 1:** Each class will receive a score equal to the inverse of their position in the rank of each ballot; **(ii) Starting at 0:** Each class will receive a score equal to the number of classes below it in the ballot rank. This method penalizes the candidate in the last position on a ballot, as it will not receive any points from this voter; **(iii) Dowdall:** In this method, each class will receive $\frac{1}{p}$ points in each ballot, where p is its position in that rank. This method favors classes that have received many first preferences compared to the other two previous methods.

4 Experimental Evaluation

4.1 Setup

The experimental environment used here is an Intel Core i5 1.8 GHz with 8 GB DDR. The ReW and CReW implementations used here are available, along with

Table 1: Borda count's policies: scores received by each candidate from the ballot of a voter; Form - Formula

Rank	Class	Starting at 1		Starting at 0		Dowdall	
		For	Score	Form	Score	Form	Score
1st	Sam	p	3	$p-1$	2	$\frac{1}{p-2}$	1.0
2rd	Bilbo	$p-1$	2	$p-2$	1	$\frac{1}{p-1}$	0.5
3rd	Frodo	$p-2$	1	$p-3$	0	$\frac{1}{p}$	0.33

other weightless models, in the C++/Python wisardpkg library[12]. The ensemble implementations are in Python. All datasets were previously preprocessed in other environments and their binarizations were serialized, so the training and test times listed here do not include the preprocessing time. The following datasets were used in the experiments: **(i)MNIST[9] and (ii) Fashion MNIST[10]:** both of them have 28x28 images, 10 classes, 60000 examples in training set and 10000 in test set; **(iii) IMDB[11]:** a dataset for binary sentiment classification with a set of 25000 highly popular movie reviews for training, and 25000 for testing. There is additional unlabeled data for use as well.

In these experiments, WiSARD and ClusWiSARD were tested with all address sizes in the range [5, 31]. The maximum number of discriminators per class of ClusWiSARD was varied in the range [3, 5]. The pre-processing used for the image datasets was the local mean threshold and for the IMDB dataset it was *tf-idf*. All ensembles were tested with a composition of 10 and 20 learners.

4.2 Results

The results are displayed in Tables 2-4. Some considerations about the experiments: (i) in Fashion MNIST the best results were: Bagging and Borda Count (BC) (start at 0), (ii) in MNIST the best models were Bagging and BC (start at 1), (iii) in IMDb the best result was Boosting, (iv) in IMDb all ensembles were more accurate than individual models, (v) considering only the ensembles, the best results in terms of training time were: Fashion MNIST - BC, MNIST - Bagging, IMDb - BC, (vi) regarding the test time: Fashion MNIST - BC, MNIST - Bagging, IMDb - Bagging, (vii) this comparison of training and testing time is not entirely appropriate, since the models that make up the ensemble are random, an ensemble may have had more ClusWiSARDs and therefore slower training. Not only structure of the ensemble and its policies influence time, but also the models that composed it and the size of their address, (viii) in general, it is expected that Bagging-based ensembles will have the fastest training, as it does not have the cost of validation, and BC will have the fastest classification since their weak learners do not perform bleaching, using only the scores of the discriminators of each weak learner and making the tiebreaker externally through the policies of the ensembles, (ix) in general, all ensembles had low

[1] Clustering parameters of ClusWiSARD: *min score* and *growth interval* parameters were kept fixed at 0.1 and 100, respectively

Table 2: The best results per model in Fashion MNIST dataset; Mod - Models; Pt - percentage of training set data used in each partition; wl - weak learners; Pol - Policies; TrT - Training Time; M5 - Maximum of five discriminators; WSD - WiSARD; Clus - Employing just ClusWiSARD models; St0 - Start at 0 (Acc's std: WSD - 0.04, Clus and BC - 0.01, Bg and Boost - 0.00)

Mod	**n**	**Pt**	**wl**	**Pol**	**Accuracy**	**TrT(s)**	**Test time(s)**
WSD	24	-	-	-	0.80	**0.42 ± 0.00**	**0.72 ± 0.00**
Clus	24	-	-	M5	0.81	1.95 ± 0.07	3.19 ± 0.02
Bg	-	0.8	10	WSD	**0.83**	3.09 ± 0.05	18.66 ± 7.93
Boost	-	-	10	Clus	0.82	19.22 ± 4.45	96.34 ± 29.47
BC	-	0.8	10	St0	**0.83**	4.37 ± 0.07	26.75 ± 3.52

Table 3: The best results per model in MNIST dataset; St1 - Start at 1 (Acc's std: all models - 0.00)

Mod	**n**	**Pt**	**wl**	**Pol**	**Acc**	**TrT(s)**	**Test time(s)**
WSD	30	-	-	-	0.89	**0.61 ± 0.13**	**0.63 ± 0.00**
Clus	28	-	-	M3	0.89	0.88 ± 0.01	2.24 ± 0.034
Bg	-	0.6	10	WSD	**0.93**	1.33 ± 0.05	10.36 ± 3.85
Boost	-	-	10	WSD	0.91	20.50 ± 5.77	57.69 ± 13.04
BC	-	0.6	20	St1	**0.93**	4.50 ± 0.09	37.61 ± 11.32

standard deviation and variance both in accuracy, training and testing times. The exception is due to the IMDb, (x) in Fashion MNIST, the best ensemble outperforms WiSARD in 3% and ClusWiSARD in 2%, (xi) in MNIST, the best ensemble outperforms both of them in 4%, (xii) in IMDb, the best ensemble outperforms both of them in 18%.

5 Conclusion and Ongoing Works

Two ensembles were constructed utilizing WiSARD and ClusWiSARD methodologies, employing established Bagging and Boosting techniques. Additionally, Borda Count ensembles were formulated based on the structural foundation of

Table 4: The best results per model in IMDb dataset; Mix - WSD and Clus; Dwd - Dowdall (Acc's std: BC - 0.01, another models - 0.00)

Mod	**n**	**Pt**	**wl**	**Pol**	**Acc**	**TrT(s)**	**Test time(s)**
WSD	5	-	-	-	0.59	**1.49 ± 0.44**	**7.33 ± 0.28**
Clus	5	-	-	M4	0.59	4.94 ± 4.94	41.94 ± 5.68
Bg	-	0.6	20	Mix	0.70	51.21 ± 5.78	22.39 ± 0.82
Boost	-	-	20	WSD	**0.77**	21.08 ± 2.52	1329.95 ± 87.27
BC	-	0.6	20	Dwd	0.70	17.71 ± 1.92	190.16 ± 33.60

WiSARD Bagging, wherein traditional voting systems are employed, and learners generate ballots through their discriminators' scores. Subsequently, these ensembles underwent evaluation across three distinct datasets, and their performance was juxtaposed with that of individual models. Notably, instances arose where WiSARD and ClusWiSARD exhibited superior performance compared to ensembles due to the absence of a pruning technique, leading to error propagation. However, in the majority of scenarios, ensembles demonstrated enhanced performance, particularly in the IMDb dataset preprocessed using *tf-idf*. The triumphant ensemble emerged as WiSARD Boosting, exhibiting an 18% higher accuracy rate than the most optimal WiSARD and ClusWiSARD configuration, achieved with a mere 20 WiSARDs. It is pivotal to underscore that no meta-learning technique was employed in the model or parameter selection process. **Ongoing works:** adding pruning policies to the ensembles, combining different preprocesses in the same weak learner, using the traditional boosting policy and use the proposed ensemble methodologies with other types of models.

References

[1] Opitz, David, and Richard Maclin, *Popular ensemble methods: An empirical study*, Journal of Artificial Intelligence Research, vol. 11, pp. 169-198, 1999.

[2] Barbosa, R., Cardoso, D. O., Carvalho, D., and França, F. M., *Weightless neuro-symbolic GPS trajectory classification*, Neurocomputing, vol. 298, pp. 100-108, 2018.

[3] da Silva Moreira, Rodrigo, and Nelson Francisco Favilla Ebecken, *Maritime Vessel Tracking with an Ensemble of WiSARD Classifiers in Video*, International Journal of Systems Applications, Engineering and Development, vol. 9, 2015.

[4] L. Breiman, *Bagging predictors*, Machine Learning, vol. 24(2), pp. 123-140, 1996.

[5] R. E. Schapire, *The strength of weak learnability*, Machine Learning, vol. 5(2), pp. 197-227, 1990.

[6] D. Lippman, *Voting theory*, Creative Commons BY-SA (A summary document), 2013.

[7] I. Aleksander, W. Thomas, and P. Bowden, *WISARD, a radical new step forward in image recognition*, Sensor Rev., vol. 4(3), pp. 120-124, 1984.

[8] Cardoso, D. O., Carvalho, D. S., Alves, D. S., Souza, D. F., Carneiro, H. C., Pedreira, C. E., Lima, P. M. V., and França, F. M., *Financial credit analysis via a clustering weightless neural classifier*, Neurocomputing, vol. 183, pp. 70-78, 2016.

[9] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P, *Gradient-based learning applied to document recognition*, Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[10] Xiao, H., Rasul, K., and Vollgraf, R., *Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms*, arXiv preprint arXiv: 1708.07747, 2017.

[11] Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., and Potts, C., *Learning word vector for sentiment analysis*, in Proceedings of the 49th AMACL, pp. 142-150, ACL, 2011.

[12] Lima Filho, A. S., Guarisa, G. P., Lusquino Filho, L. A., Oliveira, L. F., França, F. M., and Lima, P., *wisardpkg–A library for WiSARD-based models*, arXiv e-prints, arXiv-2005, 2020.

[13] Leopoldo A. D. Lusquino Filho, *Extending WiSARD to Perform Ensemble Learning, Regression, Multi-label, and Multi-modal Tasks*, PhD thesis, Federal University of Rio de Janeiro, 2021.

[14] Lusquino Filho, L. A., Oliveira, L. F., Lima Filho, A., Guarisa, G. P., Felix, L. M., Lima, P. M., and França, F. M, *Extending the weightless WiSARD classifier for regression*, Neurocomputing, vol. 416, pp. 280-291, 2020.