ISSN 2447-4924 (Impressa)
ISSN 2447-598X (Digital)

THE ROBUSTNESS OF NEURAL NETWORKS IN PATTERN RECOGNITION TASKS USING NEW TARGETS VECTORS

*José Ricardo Gonçalves Manzan[1]; Keiji Yamanaka[2]; Tiago Elias Carvalho Oliveira[3]; Igor Santos Peretta[4]; Shigueo Nomura[5]; Ana Paula Arantes Lima Manzan[6]

ABSTRACT: This paper proposes the use of new target vectors for multilayer perceptron (MLP) artificial neural networks (ANNs) in order to provide greater robustness in face of training parameter changes. These are targets with an amplified euclidean distance called orthogonal bipolar vectors (OBVs). Because of the geometric characteristics of bipolarity and perpendicularity, these targets are located in the n-dimensional space, which is the greatest possible distance from one another. This greater mutual distance of the output space points facilitates the pattern classification task of ANNs. This ensures the better performance of MLPs even in situations in which the parameters are not good for ANNs trained with conventional targets. Thus, the robustness obtained through the use of OBVs facilitates the use of MLPs for people who do not have much experience in choosing training parameters. The robustness analysis was performed using experiments for the recognition, with MLPs, of three kinds of data sets, using both OBVs and conventional targets. Real data sets used in the experiments are available at: (a) the Semeion Handwritten Digits from the Machine Learning Repository; (b) the Iris Image Database from the Chinese Academy of Sciences - CASIA; and (c) Australian Sign Language, signs of the Machine Learning Repository. The experimental results show that the use of OBVs as targets of MLPs reduces the loss of performance caused by the change of parameters. The average performance obtained with the use of OBVs is at least 15% higher than that obtained with conventional targets.

Keywords: Euclidean Distance. Multilayer Perceptron. Orthogonal Bipolar Vectors.

A ROBUSTEZ DE REDES NEURAIS EM TAREFAS DE RECONHECIMENTO DE PADRÕES USANDO NOVOS VETORES ALVO

RESUMO: Este trabalho propõe o uso de novos vetores alvo em redes neurais artificiais (ANNs) do tipo multilayer perceptron (MLP) a fim de proporcionar maior robustez diante das mudanças dos parâmetros de treinamento. Estes são alvos com distância euclidiana aumentada denominados como vetores bipolares ortogonais (OBVs). Pela característica geométrica de bipolaridade e perpendicularidade, estes alvos são localizados no espaço n-dimensional, estando a maior distância possível um do outro. Esta maior distância mútua dos pontos do espaço de saída facilita a tarefa das ANNs na classificação de padrões. Isto garante maior desempenho para MLP até mesmo em situações em que os parâmetros não são bons para ANNs treinadas com alvos convencionais. Assim, a robustez obtida por meio do uso de OBVs facilita o use de MLPs por pessoas que não tem experiência na escolha dos parâmetros de treinamento. A Análise de robustez foi realizada com a utilização de experimentos de reconhecimento, por MLPs, de três tipos de conjuntos de dados: (a) Dígitos manuscritos do Machine Learning Repository; (b) De imagens de Iris humana da Chinese Academy of Sciences – CASIA; e (c) Signos de linguagem australiana, sinais do Machine Learning Repository. Os resultados experimentais mostram que o uso de OBVs como alvos de MLPs reduz a perda de desempenho causada pela escolha de parâmetros. A média de desempenho obtida com o uso de OBVs é de até 15% maior que aquela obtida com vetores convencionais.

Palavras-chave: Multilayer Perceptron. Vetores Bipolares Ortogonais. Distância Euclidiana.

* Autor correspondente: josericardo@iftm.edu.br

1 Professor Dr. do Instituto Federal de Educação, Ciência e Tecnologia do Triângulo Mineiro - IFTM, Uberaba, MG, Brasil. josericardo@iftm.edu.br.
2 Professor Dr. da Universidade Federal de Uberlândia – UFU, Uberlândia, MG, Brasil. keiji@ufu.br
3 Bacharel em Engenharia de Computação pelo Instituto Federal de Educação, Ciência e Tecnologia do Triângulo Mineiro – IFTM, Uberaba, MG, Brasil. tiagoecoliveira@hotmail.com
4 Professor Dr. da Universidade Federal de Uberlândia – UFU, Uberlândia, MG, Brasil. iperetta@ufu.br
5 Professor Dr. da Universidade Federal de Uberlândia – UFU, Uberlândia, MG, Brasil. shigueonomura@ufu.br
6 Professora Ms. da Faculdade de Talentos Humanos – FACTHUS, Uberaba, MG, Brasil. anapaula.daramina@gmail.com

INTRODUCTION

The field of pattern recognition has gained increasing prominence in various areas of research and application. Among the various types of tools used in pattern recognition tasks, we can highlight artificial neural networks (ANNs). Among the existing ANNs, MLPs are widely used in Pattern Recognition problems. However, some studies indicate that MLPs have at least two drawbacks. Some problems, such as image recognition with medium and high resolution, or signal classification with a large amount of characteristics data with a large number of classes, require a large computational effort of MLPs (LEE et al., 2012). The other drawback concerns the definition of parameters, such as the learning rate, the size of the hidden layer and the initial synaptic weights (LAWRENCE et al., 1998). These parameters are important for a good or bad performance of MLP networks. Many studies point to methodologies for the definition of parameters (ISA et al., 2011; SIVARAM et al., 2012; SAMAL et al., 2015), but there is no exact method that leads to a good choice of parameters.

Several studies have been conducted to obtain performance gains in MLPs (HUANG et al., 2015) and to overcome the drawbacks mentioned above. There are studies dedicated to determining the topology of MLPs (SAMAL et al., 2015) and they are based on the detection of signal sensitivity to prevent noise interference concerning the performance of MLPs (LEE et al., 2006). In addition to these studies, any new proposal should highlight the addition of the sparsity regularization term for the cross-entropy cost and the update of the network parameters to minimize the joint cost (SIVARAM et al., 2012), as well as the implementation of hybrid systems, such as the use of RBF functions (ISA et al., 2011).

A detailed study of the strategies used for handwritten digit recognition competitions based on MLPs revealed that performance gains were obtained through an increase of hidden layers, an increase in the number of neurons in the hidden layers, an increase of deformed training samples to avoid the problem of "overfitting" and the use of a graphics processing unit to increase the computational processing speed (CIRESAN et al., 2012).

Only a limited number of studies have been conducted concerning the influence of the output space on the performance of MLP networks. The particular research papers found by the authors in this regard are quite old. One of these studies deals with the use of different activation functions in the output layer (RUCK et al., 1990). In this study, the behavior of linear and nonlinear functions is evaluated for the delimitation of borders. In another study, the gradient is used to obtain a threshold of the linear discriminant function of the output layer (HWANG et al., 1991). Each input pattern is a linear discriminant. Current studies focus their attention on the processing of input data, the training algorithm, the optimization of training parameters, and on the topology and hybridization incorporated with other systems.

This work proposes the use of new target vectors in MLPs for pattern recognition tasks. These targets have a geometric property of being mutually orthogonal. By being mutually orthogonal, the Euclidean distance reaches its maximum value. As a result, there is a reduction in the intersection between the convergence regions formed around each target.

With respect to the output space, preliminary studies undertaken by the authors have shown performance gains by increasing the distance between the target points. The increase of the Euclidean distance is obtained by using Orthogonal Bipolar Vectors (OBVs) (MANZAN et al., 2016). These studies show the improvement in the classification of complex problems with highly degraded patterns and with a high degree of noise. Since they are orthogonal and bipolar, the vectors have the largest possible Euclidean distance between each other.In order to show the benefits of using the OBVs proposed in this paper, the performance of the MLP was analyzed focusing on the variation of the learning rate parameter and the number of neurons in the hidden layer. This work seeks to evaluate the robustness of MLP networks trained with different targets when training parameters are changed. Through this paper, the authors show that the use of OBVs reduces the interference of the choice of training parameters on pattern recognition tasks. The data used in the experiment were: (a) the Semeion Handwritten Digits from the Machine Learning Repository, an international repository (LICHMAN, 2013); (b) the Iris Image Database from the Chinese Academy of Sciences – (CASIA, 2010); and (c) Australian Sign Language, signs of the Machine Learning Repository, international repository (KADOUS, 2002).

Section 2 presents the mathematical foundations involved. In section 3, the experimental procedure is described. The experimental results and discussion are presented in section 4. Finally, the conclusion is presented in section 5.

MATHEMATICAL FOUNDATIONS

Inner product, norm, angle, perpendicularity and euclidean distance

From a geometrical point of view, targets are vectors of n-dimensional space, so in this subsection some mathematical concepts related to the use of new targets will be presented. These concepts are the inner product, norm, angle, perpendicularity and the Euclidean distance between targets.

Let's assume that $\overline{V}$ and $\overline{W}$ are two targets of finite dimensional spaces defined by Equation (1) and Equation (2). Equation (3) represents the inner product between $\overline{V}$ and $\overline{W}$; the Euclidean distance is calculated using Equation (4); the norm of a vector is calculated by Equation (5); and Equation (6) shows the formula for calculating the angle θ between two vectors.

$$\vec{V} = \langle v_1, v_2, \ldots, v_n \rangle \quad (1)$$

$$\vec{W} = \langle w_1, w_2, \ldots, w_n \rangle \quad (2)$$

$$\langle \vec{V}, \vec{W} \rangle = v_1 \cdot w_1 + v_2 \cdot w_2 + \ldots + v_n \cdot w_n \quad (3)$$

$$d_{V,W} = \sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2 + \ldots + (w_n - v_n)^2} \quad (4)$$

$$\|V\| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2} \quad (5)$$

$$\theta = \arccos\left(\frac{\langle V, W \rangle}{\|V\|\|W\|}\right) \quad (6)$$

Two vectors are mutually orthogonal when their inner product is equal to zero.

Definition of target vectors

Conventionally, MLP networks use two types of targets in pattern recognition problems (HAYKIN, 2009). The proposal of this work it is to compare conventional targets with targets that have the characteristic of having an amplified Euclidean distance. One characteristic of these targets with amplified Euclidean distance is that they have a size that is always equivalent to a power of 2 (FAUSETT, 1994). This means that in some applications, the OBVs have a greater dimensions than conventional targets. For the sake of comparison, this work also used targets with the same characteristics as conventional targets, but with the size of OBVs. To make this information clearer, the definitions of the types of targets are presented below.

- Binary Vectors (BVs): BVs are targets consisting of n components. The value n corresponds to the number of patterns to be classified by the ANN (HAYKIN, 2009). Each row i of this matrix corresponds to the i-th BV containing the component "1" for $i = j$ and the component "0" for all other elements. Equation (7) defines a matrix with n (BVs) of size n;

$$\vec{V_{ij}} = \begin{cases} 1 \text{ for } i = j \\ 0 \text{ for } i \neq j \end{cases} \quad (7)$$

- Conventional Bipolar Vectors (CBVs): In a similar way to what happens with BVs, CBVs are also composed of n components and their size depends on the number of patterns to be classified by the ANN (HAYKIN, 2009). Equation (8) defines a matrix with n (CBV)s of size n;

$$\vec{V_{ij}} = \begin{cases} 1 \text{ for } i = j \\ -1 \text{ for } i \neq j \end{cases} \quad (8)$$

- Orthogonal Bipolar Vectors (OBVs): These targets are mutually perpendicular. The algorithm that allows OBVs to be obtained is presented in Fausset, 1994. A feature of these targets is that half of the elements are equal to "1" and the other half is equal to "–1". For mathematical reasons, the size of the OBV is always a power of "2";
- Non-Orthogonal Bipolar Vectors (NOV): These targets are an extension of CBVs. They have the same characteristics of the CBV-type targets and the same size as OBVs. They were used in this work only in order to provide a fair comparison between targets with the same characteristics as CBVs but having the same size as OBVs. To obtain NOVs, CBVs are therefore complemented with the term "–1" in order to achieve the same size as the OBVs.

Observations about the target

According to their definition, the elements of the CBVs are equal to –1 except for one element whose value is 1. In addition, the difference between these targets is given by the position occupied by element 1. Consequently, according to the inner product definition given by Equation (3), the result of the inner product can be determined between two CBV networks of sizes greater than 4, as a function of the number of components. For the indexes where the component of the vectors is equal to 1, the products of the components are equal to –1, and for other indexes, the products of the components are equal to 1. Therefore, two components for the products will be equal to –1 and all others will be equal to 1. This brings us to the value of the inner product between CBVs as a function of the number n of components, as shown in Equation (9).

$$\langle \vec{V}, \vec{W} \rangle = n - 4 \quad (9)$$

Based on the norm definition given by Equation (5), one can also express norm of the vectors for CBVs as a function of the number of components, according to Equation (10).

$$\|V\| = \sqrt{n} \quad (10)$$

Substituting the results of Equations (9)-(10) in Equation (6), which corresponds to the angle between vectors, we get Equation (11), which expresses the value of the arccosine of the angle between these vectors as a function of the number of components of the CBVs.

$$\theta = \arccos\left(\frac{n-4}{n}\right) \quad (11)$$

If we increase the value of n as much as possible, the ratio $(n - 4)/n$ will tend to 1. The arccosine function decreases and approaches 0 when its argument tends to 1. Thus, the inner product between two CBVs increases as their size increases, decreasing the angle between them. In other words, if the dimensionality of the output space is high enough, the difference between two CBVs will become increasingly ambiguous. Furthermore, when taking two CBVs of any given size, the Euclidean distance shown in Equation (4), is always equal to $2\sqrt{2}$.

On the other hand, if the inner product between two vectors is zero, then the numerator, the ratio established in Equation (6), is zero. Thus, the angle between both vectors is 90 degrees. In this case, the vectors are perpendicular. Equation (12) shows the Euclidean distance between two OBVs for the number of components n.

$$d_{V,W} = \sqrt{2n} \tag{12}$$

This shows that the Euclidean distance between OBVs is greater than the Euclidean distance between CBVs if the number of components is greater than 4. It also shows that the Euclidean distance between the OBVs increases as you increase the number of components. The greater the size of the OBV-type targets, the greater the Euclidean distance between them.

BV-type targets are mutually perpendicular. However, the Euclidean distance between each pair of BVs is always equal to $\sqrt{2}$

NOVs may have the same size as OBVs, but their inner product is not-zero. The difference of NOVs in relation to CBVs is only the size. The inner product between two NOV-type targets is also always equal to . If the size of the NOVs is large, then the corresponding inner product between them is large. We therefore assume that the increase in the NOV's size does not improve performance in MLPs (MANZAN et al., 2016).

Targets used in the experiments

The dimensions of the targets are directly linked to the types of data that were used in this work. This work used three types of benchmark data. Handwritten digits 0 to 9, human iris images and signs corresponding to the Australian sign language, which we call Auslan Signs. A detailed description of this data is given in section 3.1.

Ten classes need to be classified in the experiments with handwritten digits. As such, CBVs of size 10 (CBV10) are sufficients for solving the problem. As established in subsection 2.2, OBV-type targets have a size equal to a power of 2. To classify 10 digits, OBVs with size 16 (OBV16) must be used. To ensure that comparisons were fair, NOVs of size 16 (NOV16) were used.

The data corresponded to the irises of 71 individuals in the experiments using the human iris. Following the same reasoning used with handwritten digits, CBVs of size 71 (CBV71), and NOVs (NOV128) and OBVs (OBV128) of size 128 were used. Finally, since there are 95 Auslan signs, CBVs of size 95 (CBV95) and NOVs (NOV128) and OBVs (OBV128) of size 128 were used.

EXPERIMENTAL PROCEDURE

Database

This work was performed with three types of benchmark data. The reason for the use of three types of data is to assure with the highest possible reliability that the methodology is superior to the traditional MLP approach, even with data of different nature and with different classes. By showing that the proposed method is more robust and reliable than the traditional approach, a potential MLP user in a pattern recognition task will not have concerns about defining the network topology and the initial training parameters. In addition, it allows the user to perform his experiments with a reduced topology size and, consequently, in a shorter time.

a. Handwritten Digits

Digits obtained from the international repository known as the Semeion Handwritten Digits of Machine Learning Repository were used in the training of the MLP networks (LICHMAN, 2013). These patterns were obtained from a group of about 80 individuals, who were asked to write down the digits from 0 to 9 twice. In the first request, people were asked to write the digits calmly, prioritizing perfection in writing. In the second request, people were asked to write the digits quickly, without worrying about their readability.

Each figure was scanned in an image containing 256 pixels in the format of 16 rows and 16 columns. Each image was processed in a resolution scale of 256 gray levels. The pixel matrix was subsequently transformed into a line vector of 256 components, with each line being positioned immediately to the right of its top line in the matrix. Each pixel corresponding to the background of the image was assigned a value of 0 and each pixel corresponding to the digit was assigned the value 1 according to the information described in the repository (LICHMAN, 2013). In this work, the pixels corresponding to the image background received the value –1 instead of 0.

b. Iris Image Database

The authors also conducted experiments with human irises obtained from the Chinese Academy of Sciences - Institute of Automation database called CASIA, 2010. The database contains iris images from 108 subjects, 71 of which consisting of complete data with six images. For this reason, we adopted the data corresponding to these 71 subjects. For each test subject four images were used in the training session. According to the CASIA repository, these images were taken by using infrared light to obtain the iris features with enough contrast for biometric pattern recognition.

The steps for iris image processing are as follows. The first step is the locate the iris region in the image, which is done using the circle Hough transform. Subsequently, the ring-shaped iris region is normalized in order to be represented it as a rectangular matrix. Finally, the extraction of the iris features is performed. In this paper, the iris features were extracted by convoluting the normalized image with the so-called log Gabor filter. Filtering gives rise to complex coefficients, whose phases are quantized to one of the four quadrants of the complex plane. Each quadrant is referenced by two bits, and a binary template is created (NEGIN et al., 2000). For each image there are 8640 pixels arranged in 18 concentric circles, each containing 480 pixels.

In order to reduce the computational effort without significant loss of performance, the average for every 10 pixels of the circumference was calculated. As such, each circle was assigned 48 values corresponding to all averages. Thus, 240 values were allocated to each pattern.

c. Australian Sign Language Signs

In addition to the data presented above, experiments were performed with Auslan signs (Australian Sign Language). The authors captured 27 samples for each of the 95 Auslan signs using high-quality position trackers from native individuals (KADOUS, 2002).

Each sample is represented by 60 rows with 22 values. The 22 values correspond to 22 data capture channels (KADOUS, 2002). Only the first four rows were used of each sample from the analysis of preliminary experiments in order to reduce the computational effort.

Experimental and statistical design

To achieve the objectives of this work, training and testing experiments were conducted in MLPs for the recognition of handwritten digits, human irises and Auslan signs. For each type of data, experiments were performed with 7 different values for the initial learning rate and 12 different values for the number of neurons in the hidden layer. All combinations of initial learning rates and number of neurons of the hidden layer were subjected to the use of three types of characteristic targets for each data set: CBV10, NOV16 and OBV16 for handwritten digits, CBV71, NOV128 and OBV128 for human irises and CBV95, NOV128 and OBV128 for Auslan signs.

The experiments were repeated 100 times for each combination, with random initialization of weights. The initial synaptic weights were random values between –0.5 and 0.5. The initial learning rate values adopted for handwritten digits were: 0.001, 0.005, 0.01, 0.05, 0.1, 0.25 and 0.4. The initial learning rate values adopted for human irises and Auslan signs were: 0.0001, 0.00025, 0.0005, 0.00075, 0.001, 0.0025 and 0.005. The values for the number of neurons in the hidden layer were: 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1000. As such, each type of target was subjected to 84 combinations of initial learning rate and number of neurons in the hidden layer for each data type. Since 100 experiments were performed for each combination, each type of target was subjected to 8400 experiments. Thus 25200 experiments were therefore performed for each of the three targets types: CBV, NOV and OBV. Finally, 75600 experiments were conducted accounting for the three types of data used.

Each training session lasted 50 epochs. For each epoch, a distinct set of tests in relation to the training set was submitted to the network, and the hit rate (accuracy) was collected. Each experiment brought in 50 sets of performance results corresponding to each epoch, therefore. The mean and variation coefficient for the 25200 experiments performed for each type of target were calculated. Box-plot graphics were also generated for epochs 1, 10, 20, 30, 40 and 50. Those parameter combinations (out of a total 84 combinations) with a performance rate of over 70% in epochs 1, 10, 20, 30, 40 and 50 were written down.

A total of 90 samples of each digit were used for the MLP network training step, reaching a total of 900 samples for the 10 basic handwritten digits. Two samples of each individual were used for the training step with the human iris. As such, 142 iris samples were used. Nine samples of each sign were used for Auslan the signs, that is, a total 855 samples. The test sets for handwritten digits, human irises and Auslan signs were composed of 450, 213 and 855 samples, respectively.

The experiments were performed with computers using exactly the same operating system settings. The authors created the simulation program using the Matlab® 2013 software. The algorithm used the adaptive learning rate and momentum term.

EXPERIMENTAL RESULTS AND DISCUSSION

Figures (1)-(3) show the mean of performance (accuracy) from all parameter combinations for each of the 50 training epochs. In Figure (1), the results for the experiments with handwritten digits are given. The results for the experiments with human irises and Auslan signs are shown in Figures (2) and (3), respectively.

The mean of performance obtained with the use of OBVs is higher in all 50 epochs of the various simulations combining different parameters. However, this difference is much clearer in the early training epochs. This shows that the use of OBVs provides good performance with little computational effort as opposed to conventional targets that require more training.

Another interesting phenomenon is that the average performance obtained with OBVs grows with increasing epochs without oscillations. With conventional targets, one can see that oscillations occur in the experiments with handwritten digits with the increase of epochs. This shows that the convetional targets make the network more susceptible to the overfitting effect than when OBVs are used.

Figure 1: Mean of performance mean for all epochs – handwritten digits.

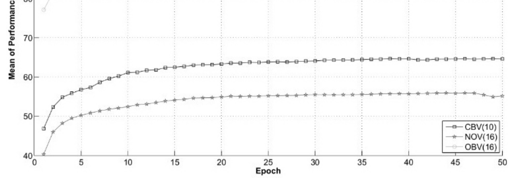
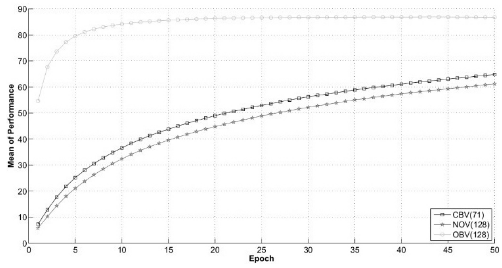


Figure 2: Mean of performance for all epochs – human irises.



Figures (4), (5) and (6) show the coefficients of variation in performances obtained from experiments with handwritten digits, human iris and Auslan signs, respectively. A significantly lower variability is observed in the experiments in which OBVs are used. This shows how much the use of OBVs improve the robustness of MLP network performance. In other words, networks trained with the use of OBVs suffer less interference from parameter changes. One can also see that the difference between variabilities in the experiments with CBVs and NOVs is very small. This shows that increasing the size of CBVs does not improve the performance of the MLP. At times, there is even an increase in variability when NOVs are used.

Figure 3: Mean of performance for all epochs – Auslan signs.

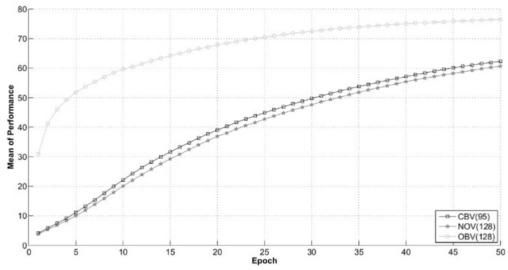


Figure 4: Coefficients of variation for the performances obtained in training – handwritten digits.

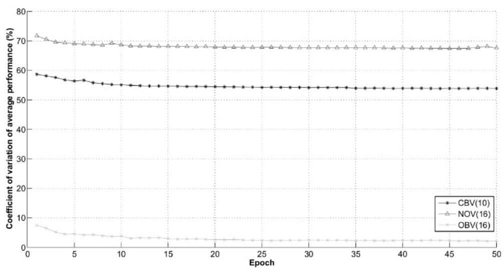


Figure 5: Coefficients of variation for the performances obtained in training – human irises.

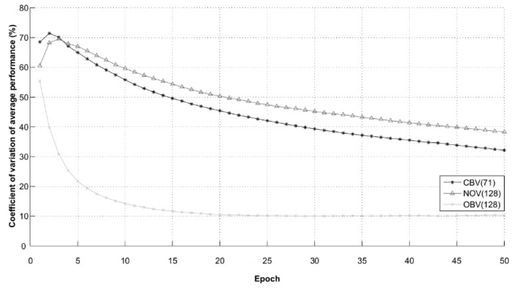
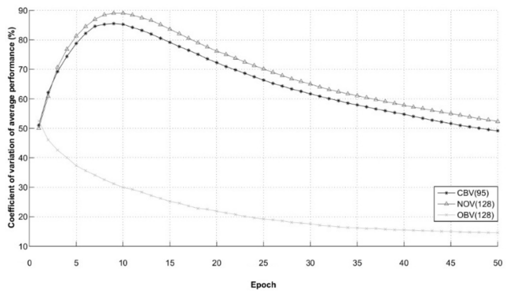


Figure 6: Coefficients of variation for the performances obtained in training – Auslan signs.



Figures (7), (8), (9), (10), (11) and (12) show the box-plot graphs relating to the epochs 1, 10, 20, 30, 40 and 50, respectively. The box-plot graphs refer to the performances obtained with all the parameter combinations.

The box-plot graph of figure 7 shows that experiments using OBVs comprise a superior range of performance than experiments with conventional targets. In the experiments with human irises and Auslan signs, the variability presented by OBVs is much higher. However, this variability comprises a superior range in performance. The variability with conventional targets is small, but shows a concentration of less than 10%. In experiments with handwritten digits, the variability is lower with OBVs and higher with CBVs and NOVs. This difference is explained by the fact that the handwritten digits have only 10 classes, whereas the human irises have 70 classes and Auslan signs have 95 classes.

Figure 7: Performance for epoch 1.

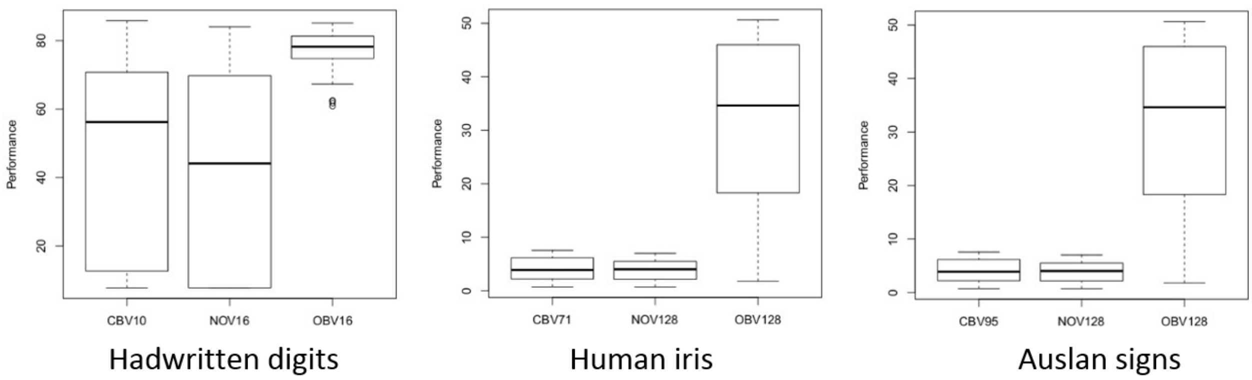
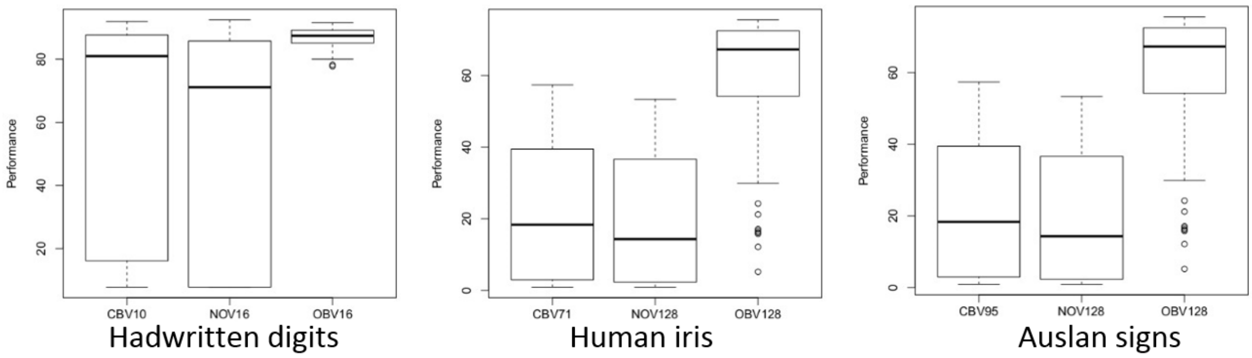
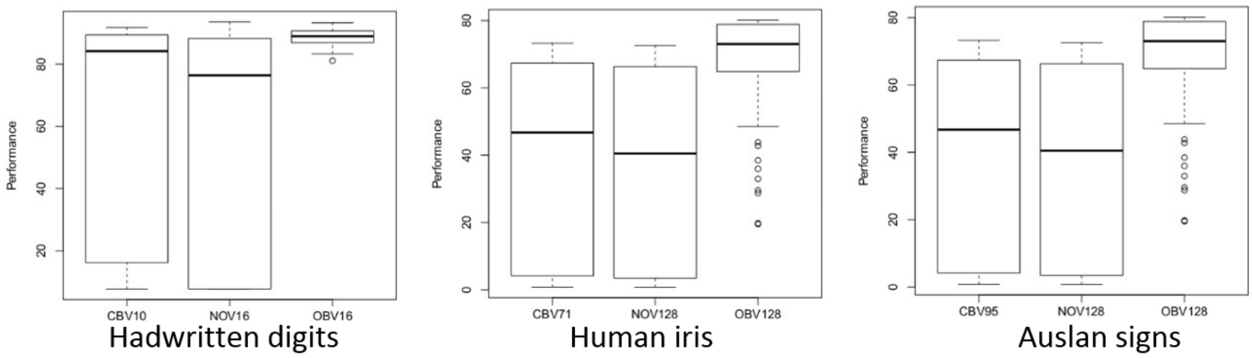


Figure 8: Performance for epoch 10.



Figures (9) - (12) show that the variability of the performances obtained in the experiments with OBVs reduces more and more with the increase of the epochs.

Figure 9: Performance for epoch 20.



In addition, the distribution of performances is concentrated in the upper range of the box-plot chart. On the other hand, the variability of the performances obtained in the experiments with conventional targets increases as the number of epochs increases. These results reveal how OBVs are robust to the effects of variations in training parameters.

Figure 10: Performance for epoch 30.

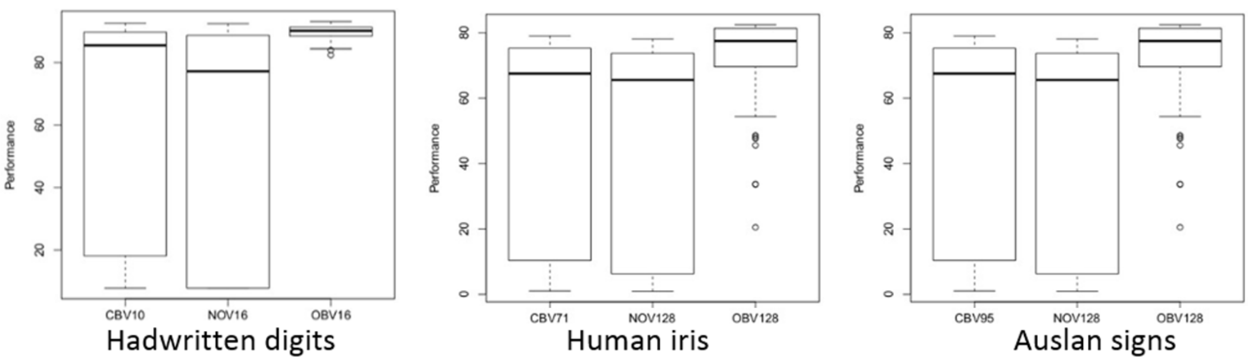


Figure 11: Performance for epoch 40.

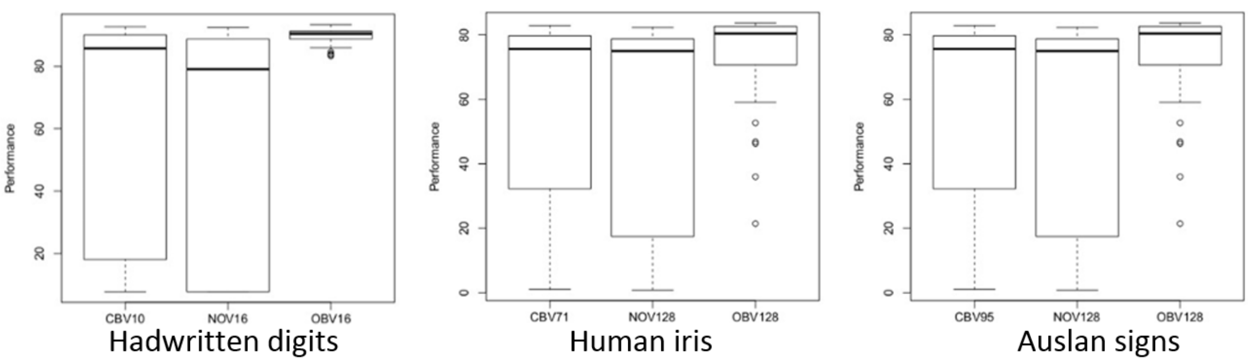


Figure 12: Performance for epoch 50.

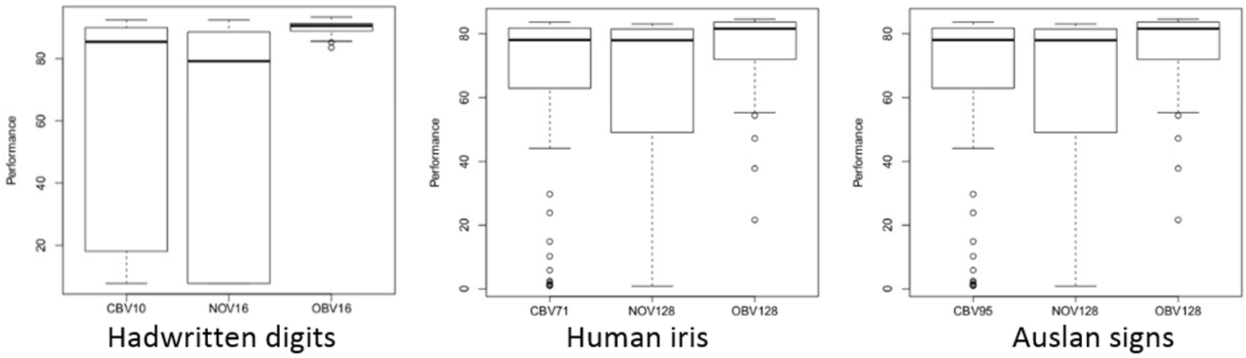


Table 1 shows the number of experiments with results higher than 70% for each type of target. The results refer to all 84 experiments formed by the combination of the parameters for handwritten digits, human irises and Auslan signs.

Table 1 - Number of experiments with superior results to 70%

Target Vectors		Handwritten digits			Human irises			Auslan signs		
		CBV 10	NOV 16	OBV 16	CBV 71	NOV 128	OBV 128	CBV 95	NOV 128	OBV 128
Epochs	1	22	21	74	0	0	40	0	0	0
	10	52	43	84	0	0	74	0	0	32
	20	55	47	84	15	9	76	14	11	52
	30	55	49	84	27	22	77	39	37	61
	40	57	49	84	38	34	77	49	48	64
	50	57	47	84	49	44	76	60	59	68

These results show how networks trained with OBVs are generally capable of obtaining high performances regardless of the datasets used and the parameters used in the training. A user with little experience in MLP networks, therefore, would not have much difficulty in modeling and training the network with good performance rates if they adopted OBV-type targets.

CONCLUSION

This paper proposes the use of OBVs as targets in MLP-type networks; they provide superior performance when compared to more conventionally used targets, such as CBVs and NOVs. It also proposes that the performance of the MLP receiving less influence from the choice of parameters of the initial learning rate and number of neurons of the hidden layer when OBVs are used. In other words, it proposes that the use of OBVs makes MLP networks more robust in face of variations in both training parameters and the types of data to be classified.

The experimental results showed that the mean performance obtained with the use of OBVs is higher than the mean performance obtained with the use of conventional targets. This phenomenon occurs independently of the type of data to be classified and the training parameters. The use of OBVs in pattern recognition problems therefore guarantees a better performance compared to the use of conventional targets.

The results also showed that changes in training parameters interfere much less in the performance of MLPs when OBVs are used. The performance variation coefficients are much lower when this type of target is used. The box-plot graphs also make clear that the dispersion of performances of trained OBV networks is very small after 10 epochs and concentrates on the upper parts, above 80%. This means that when using MLPs in pattern recognition tasks, users with little knowledge of MLP network modeling will have little concern about finding the training parameters that will provide the best performance. It also indicates that it is not necessary to use very large topologies, reducing the computational effort required in the training step. Another aspect that may contribute to the reduction of computational effort is the fact that the use of OBVs accelerates the convergence of training. With few training epochs, the MLP achieves high performances if the number of classes of the problem is small, as shown by the results obtained with handwritten digits.

This way, MLP networks become more robust when OBVs are used as targets. Such an approach is simple to implement since it does not require changes in the training algorithm, in the generation of synaptic weights and in the determination of the network topology. The only topological change is in the output layer, since depending on the number of problem classes to be solved, OBVs will have to be used that are larger than conventional targets.

REFERENCES

CHINESE ACADEMY OF SCIENCES (CASIA). **Human Iris**: database of 756: greyscale eye images. [S.l.]: Institute of Automation, 2010.

CIRESAN, D.; MEIER, U.; GAMBARDELLA, L.; SCHMIDHUBER, J. Deep big Multilayer perceptrons for digit recognition. . **Lecture Notes in Computer Science**, v. 7700, p. 581-598, 2012. DOI: 10.1007/978-3-642-35289-8_31.

FAUSSET, L. **Fundamentals of neural networks**: arquitectures, algorithms, and applications. Prentice-Hall, Inc., 1994.

HAYKIN, S. **Neural networks and learning machines**. v. 03, Pearson Education Upper Saddle River, 2009.

HUANG, G.; HUANG, G. B.; SONG, S.; YOU, K. Trends in extreme learning machines: a review. **Neural Networks**, v.61, p. 32-48, 2015. DOI: http://dx.doi.org/10.1016/j.neunet.2014.10.001.

HWANG, J. N.; CHOI, J. J.; OH, S.; MARKS, R. J. Query-based learning applied to partially trained multilayer perceptrons. **IEEE Transactions on Neural Networks**, v. 02, n. 01, p. 131-136, 1991. DOI: 10.1109/72.80299.

ISA, Nor Ashidi Mat; MAMAT, Wan Mohd Fahmi Wan. Clustered-hybrid multilayer perceptron network for pattern recognition application. **Applied Soft Computing**, v. 11, n. 1, p. 1457-1466, 2011.

KADOUS, M. W. **Temporal classification**: extending the classification paradigma to multivariate time series. New South Wales: The University of New South Wales, 2002.

LAWRENCE, S.; BURNS, I.; BACK, A.; TSOI A.; GILES, C. L. Neural network classification and prior class probabilities. **Lecture Notes in Computer Science**, v. 1524, p. 299-313, 1998. DOI: 10.1007/3-540-49430-8_15.

LEE, C. H.; CHANG, F. K.; KUO, C. T.; CHANG, H. H. A hybrid of electromagnetism-like mechanism and back-propagation algorithms for recurrent neural fuzzy systems design. **International Journal of Systems Science**, v. 43, n. 02, p. 231-247, 2012. DOI: 10.1080/00207721.2010.488758.

LEE, C. M.; YANG, S. S.; HO, C. L. Modified back-propagation algorithm applied to decision-feedback equalisation. **IEE Proceedings Vision**, Image and Signal Processing, v. 153, n. 06, p. 805-809, 2006. DOI: 10.1049/ip-vis:20050139.

LICHMAN, M. **UCI Machine Learning Repository**. Irvine: School of Information and Computer Sciences, University of California, 2013.

MANZAN, J. R. G.; YAMANAKA, K.; PERETTA, I. S.; PINTO, E. R.; OLIVEIRA, T. E. C.; NOMURA, S. A mathematical discussion concerning the performance multilayer perceptron-type artificial neural networks through use of orthogonal bipolar vectors. **Computational & Applied Mathematics**, v. 01, p. 01-12, 2006. DOI: https://doi.org/10.1007/s40314-016-0377-x

NEGIN, M.; CHMIELEWSKI; SALGANICOFF, M.; CAMUS, T. A.; VON SELEN, U.M.C.; VENETIANER, P.L.; ZHANG, G.G. An Iris Biometric System for Public and Personal Use. **Computer**, v. 33, n. 02, p. 70-75, 2000. DOI: 10.1109/2.820042.

RUCK, D. W.; ROGERS, S.K.; KABRISKY, M.; OXLEY, M.E.; SUTER, B.W. The multilayer perceptron as approximation to a Bayes optimal discriminant function. **IEEE Transactions on Neural Networks**, v. 01, n. 04, p. 296-298, 1990. DOI: 10.1109/72.80266.

SAMAL, A.; PANDA, J.; DAS, N. Performance Comparison of Single-Layer Perceptron and FLANN-Based Structure for Isolated Digit Recognition. **Advances in Intelligent Systems and Computing – Intelligent Computing, Communication and Devices**, v. 308, p. 237-246, 2015. DOI: 10.1007/978-81-322-2012-1_25.

SIVARAM, G. S. V. S.; HERMANSKY, H. Sparse Multilayer Perceptron for Phoneme Recognition. **IEEE Transactions on Audio, Speech, and Language Processing**, v. 20, n. 01, p. 23-29, 2012. DOI: 10.1109/TASL.2011.2129510.