БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ . BULGARIAN ACADEMY OF SCIENCES

КИБЕРНЕТИКА И ИНФОРМАЦИОННИ ТЕХНОЛОГИИ · Том **3**, № 1
CYBERNETICS AND INFORMATION TECHNOLOGIES · Volume **3**, No 1

София . 2003 . Sofia

Wavelets' Application for Improving the Neural Network Recognition of Text Images

Dimo Dimov, Sergei Milanov

Institute of Information Technologies, 1113 Sofia, E-mail: dtdim@iinf.bas.bg

Abstract: *The paper presents usage of wavelet transform for pre-processing of graphic objects, liable to recognition. It is investigated the possibility for preliminary structuring and reducing of the information, passed to a traditional classifier – nonlinear three-layer perceptron, to improve its recognition opportunities. Experiments performed for recognition of isolated letters are briefly reported. The analysis of the experimental results shows that optimal levels can be defined for reducing and structuring of the input information. A heuristic algorithm is also suggested for a priori evaluation of these levels, i.e. for a preliminary definition of the number of necessary neurons in the input layer of the perceptron.*

Keywords: *image pre-processing, wavelet transform, nonlinear multi-layer perceptron, structuring and reducing of features for recognition.*

1. Introduction

The recognition of isolated letters is a basic module in the computer systems for document analysis. The neural network approach is often applied for building of this module, usually called a character classifier [1, 2, 3].

Regardless of the recent advance in neural networking, there is still more to desire from neural networks, either in the current case or in other applications. Letter recognition, i.e. classification of character images, can be considered as a particular and representative case of the neural network recognition of two-dimensional (2D) images at all. In other words, the computational complexity of recognition could be manipulated to a great extent by variance of the letter number for recognition, i.e. the alphabet, by variance of letter fonts and styles, i.e. the capacity of the recognition classes, by the magnitude of image noise, etc.

On the other hand, the success of the wavelet transform (WT) application to image processing and especially to image compression is well known. As far as the compression could also be considered in terms of recognition, the WT (like Fourier transform) could be used for the task, often called feature space reduction.

The paper reports the results of WT usage for preliminary structuring and reducing of the information passed to the neural networks for recognition of isolated letters (or characters).

2. Theoretic milestones of the considered task

A recognition system can be generally represented by the scheme in Fig.1.

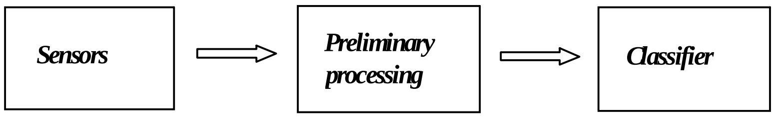

Fig. 1. A common functional scheme for recognition

Following the common practice of using ***neural networks*** as classifiers, we have chosen a multi-layer Nonlinear Perceptron (NP) of classic type [4, 5] for our investigation. The necessary training procedures are based on the well-known "back-propagating error" method. Aiming at simplicity of experiments, a ***three-layer* NP** scheme is considered hereinafter, i.e. NP with a single hidden layer of uniform neurons, besides the input and the output neuron layers.

The accent of investigation is put on the ***Preliminary Processing*** **(**PP**)**, and more precisely, on the usage of wavelet transforms for structuring and/or reducing of primary image data, before passing it to the NP. The aims are twofold – to decrease both the recognition error probability and the duration for effective training of NP.

It is popular that in comparison with the Fourier transform, the wavelet transform localizes better in both the domains – the spatial and the frequency domain [6]. A given image (of $2N\times 2N$ pixels) could be decomposed into four sub-images – one approximated subimage $A(I)$ (of $N\times N$ pixels) and three detailed subimages $D_H(I)$, $D_V(I)$ and $D_D(I)$ (of $N\times N$ pixels each). These detailed sub-images reflect the difference between the initial original image I and its approximation $A(I)$ performed by a two-dimensional WT. WTs are reversible, i.e. the four sub-images contain the same information as the origin image. A visual interpretation of WT is illustrated in Fig.2.

Next WTs can be also applied over the just obtained approximated subimage $A(I) = A^1(I)$, WT by WT, until the current approximated subimage $A^n(I)$, reaches a minimal volume, defined by the WT filter used, where n is the number of WT-subimages (WTs) applied. The pixels of the WTs are non-correlated to a high extent, because of the orthogonal basis principle used in WTs. In this way, the origin image I can be transformed into a sequence of detailed subimage triples, starting with the currently processed subimage $A^n(I)$, n>0 [6, 7, 8].

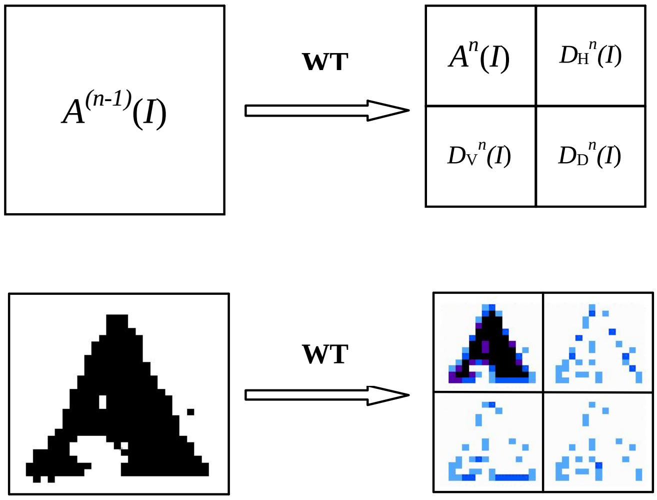

Fig. 2. Wavelet transform – a visual interpretation

The WTs can be easily implemented following the pyramidal algorithm of Mallat [8], which is based on one-dimensional convolutions of the rows and columns of $A^n(I)$ with the one-dimensional quadrature mirror filters H and G. The experiments here are based on the Daubechies' filters $db(k)$ in the range k = 1, ..., 5.

3. Description of the experiment

The behavior of three-layer NPs used for classification is investigated. The objects for classification are isolated letters of Latin alphabet, represented as 2D images.

The pixel intensities of the image are passed to the input layer of the perceptron under investigation; i.e. the number of input neurons is equal to the image size in pixels. The size of all experimental images is one and the same, equal to (32×32). Statistic values, i.e. posterior probabilities, are expected on the perceptron output for all the set of objects under recognition, i.e. the character images. The objects are considered recognized using a majoring rule for the NP-output probabilities.

The training of all perceptions used in the experiment is made with one and the same learning sequence. It includes letters of three different fonts, and the testing set includes letters of five other fonts.

After a period of training epochs, the following conditions are currently checked:

- the quality of training: by evaluation of the error probability when the perceptron is tested with objects from the training set;
- the quality of generalizing: the same as above, but for objects from the testing set.

The essence of experiment consists in the appropriated structuring and/or reducing of the entered 2D images performed by the preliminary wavelet transform.

We call ***"image structuring"*** the application of one WT on a given 2D-image, where the resulting image is arranged in accordance with the scheme of Fig.2. Obviously, the structuring doesn't change the image size. Additionally, we call ***"image reducing"*** the elimination of detailed subimages (refer to Fig. 2), where the image size decreases twice. Only the approximated subimage is going to be used after reducing.

As a ***“level of structuring”*** we consider the number of consecutive structuring operations, applied on the image, and as a ***“level of reducing”*** – the respective number of reducing operations. Obviously, both the operations are commutative, and the current level of structuring should be always greater than the one of reducing.

The LEVEL/SIZE-scheme, represented in Fig.3, shows the place of each experiment in a common logical structure. The scheme is organized like a rectangular coordinate table, in accordance with the number of structuring and reducing operations performed on the images under recognition, and its main items have the following sense:

(1) Horizontally (by LEVEL): the level of structuring increases from left to right, but the number of input neurons remains unchanged from perceptron to perceptron under investigation.

(2) Vertically (by SIZE): the level of reducing increases from top to bottom, and the input neuron number of corresponding perceptron decreases 4 times with each reduction.

(3) Each SIZE/LEVEL-cell of the table corresponds to a perceptron under investigation. The cell is darkening respectively to the image part passed to the perceptron input as pixels.

So, the goal of experiments can be defined as chosing the best experiment, or perceptron in the SIZE/LEVEL-table, in order to enhance both the training speed and the generalizing quality. We try also to formalize the results.

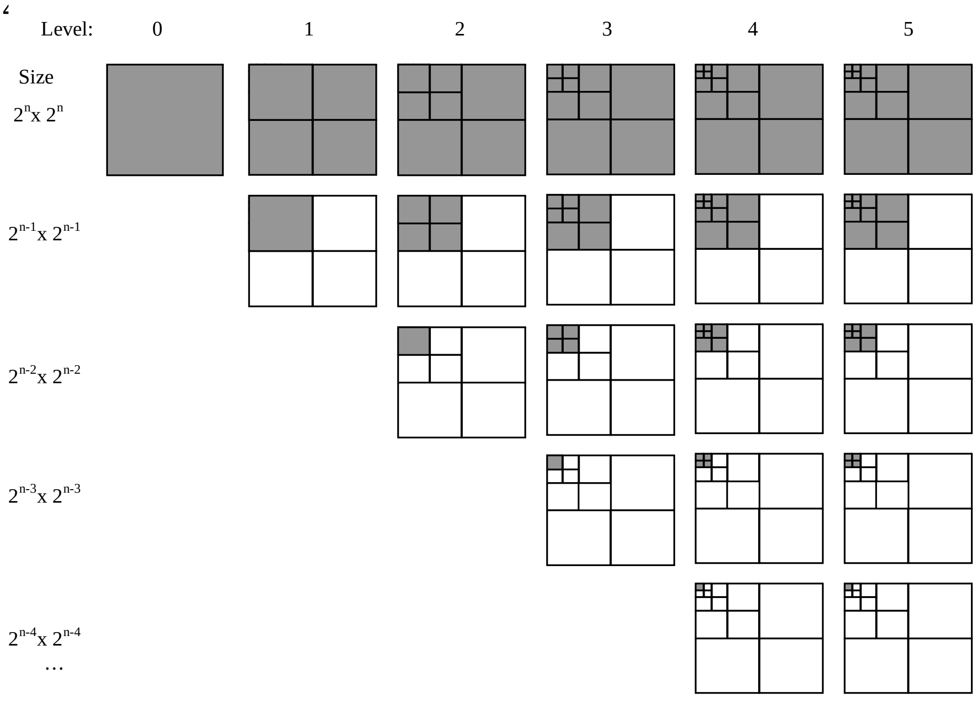

Fig. 3. The common logical structure of the experiments (SIZE/LEVEL-table)

The experiments are performed for the set of capital Latin letters. Experimental results are illustrated in Fig.4, in tables of structure similar to that of the LEVEL/SIZE-table (refer to Fig.3). The lighter a cell herein is, the smaller the error of the respective perceptron is.

The results for "negative" images, i.e. white letters on black background, are not illustrated because of being quite similar to the "positive" images' results.

5. Analysis of the experimental results

The analysis is based on the SIZE/LEVEL-table of Fig.3, and the illustrations of the experimental results in Fig.4.

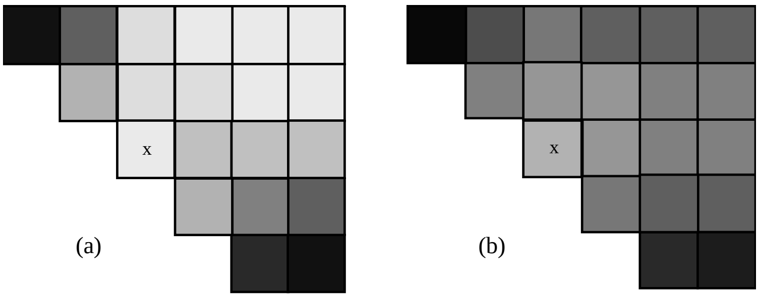

Fig. 4. Results for black letters on white background:

a) quality of training (by equal duration of training); (b) quality of generalizing

5.1. Structuring procedures – on the LEVEL-axis

In this experimental direction, the respective perceptrons received an equal information, but it was differently structured in the respective sequence of WTs.

Obviously, an "***optimal level of structuring***" (**OLS**) can be defined, on the left of which (Fig. 2) the structuring leads to improvement of the results, and on the right the results are comparatively equal. In the explored case, this **OLS** can be evaluated as equal to level 2.

5.2. Reducing procedures – on the SIZE-axis

In this experimental direction, the respective perceptrons received reduced information; more specifically – only the approximated subimages (resulting after WT-processing) are passed to the perceptron. It is considered that the detailed subimage information, or at least the one from the initial levels of reducing, is not significant for the recognition, i.e. that it mostly contains the "noise" of the original image. And if this noise is low, then the detailed subimages contain mainly negligible values (close to zero).

This supposition, known by the theory of WT and Fourier analysis as well [6, 8] is shown in the results represented in Fig.4.

Moreover, a deterioration of the results was observed reducing the image below a level, that we call "***Optimal Level of Reducing*** " (OLR), in our case, OLR = 2. The latter could be explained by the fact that after OLR the detailed subimages contain more significant information for the original image.

5.3. An optimal construction for the perceptron input

The SIZE/LEVEL-table of the experiments (refer to Fig.3) is related only to the number of structuring and reducing operations made over the origin image. Besides, there isn't any need to arrange the operations in order to approach the respective perceptron in the table. So, an "***Optimal Structure of the Perceptron Input***" (OSPI) can be intuitively defined, using the cell-position of coordinates (OLR, OLS) in the table. It is obvious, that

$$0 \le \mathrm{OLR} \le \mathrm{OLS} \le \min\{\log_2 N, \log_2 M\},$$

where N and M are the sizes of the origin image (for each object under recognition). So, we can state for a definite class of objects for recognition:

- In accordance with the OSPI, the size of the input layer of the perceptron (in number of neurons) equals $NM/4^{(\mathrm{OLR})}$.
- Subimages to be passed to the perceptron have sizes equal to $(N/2^{(\mathrm{OLR})}) \times (M/2^{(\mathrm{OLR})})$. They can be extracted from the origin image by a number of (OLS-OLR) consecutive structuring operations and by a number of OLR structuring and reducing operations.

5.4. Heuristic algorithms for a priori appointment of the optimal levels

A low boundary can be fixed for the OLR-value in the case of recognition of the graphic objects (as letters, numbers, etc.), via the following heuristic ***algorithm A***:

$\langle$A0$\rangle$ Binarize all the images for recognition.

$\langle$A1$\rangle$ Scan each image horizontally and compute the number α_i of all pixel transitions of type "0-1" and "1-0", for each scan row i of the image. Find the greatest value α, $\alpha = \max(\alpha_i)$, $0 \le i \le N$, where N is the horizontal size of the image in pixels.

$\langle$A2$\rangle$ By analogy, scan the image vertically and compute the respective maximal value β, $\beta = \max(\beta_i)$, $0 \le i \le M$, where M is the vertical size in pixels.

$\langle$A3$\rangle$ Compute the value γ, i.e. a low boundary for OLR, using the formulae:

$$\gamma = \left\lceil \min\left\{ \log_2\left(\frac{N}{\alpha+1}\right), \log_2\left(\frac{M}{\beta+1}\right) \right\} \right\rceil .$$

So, in the current case of classification (of letters and digits), the level OLR is equal to 2, i.e. in the SIZE-direction it is recommended to reduce the images to size 8x8.

For a preliminary appointment of the OLS-value, the following heuristic ***algorithm B*** can be suggested:

$\langle$B0$\rangle$ Perform a sequence of WTs for structuring the images in the training set, while the following inequality is true:

$$\xi_D \le \varepsilon\, \xi_A ,$$

where ξ_A is the average intensity of the approximated subimage $A^{n+1}(I)$, and ξ_D is the average intensity of the three detailed subimages $D_{\mathrm{H}}^{n+1}(I)$, $D_{\mathrm{V}}^{n+1}(I)$ and $D_{\mathrm{D}}^{n+1}(I)$. The constant ε could be expertly defined in the interval (0.1, 0.2).

$\langle$B1$\rangle$ Let the number of performed WTs be equal to μ. Then, if $\mu \le \gamma$, where γ is computed in the algorithm $\boldsymbol{A}$, we accept OLS = γ; otherwise, we accept OLS = μ.

If one of the two algorithms doesn't work properly, due to some reasons, it is recommended to extend the obtained more plausible result applying the following equality:

$$\text{OLS} = \text{OLR}.$$

6. Conclusion

It is investigated and is experimentally proved that the recognition plausibility of multilayer nonlinear perceptron is improving by a preliminary WT-structuring and WT-reducing of the information passed to its input layer, following the table suggested herein (see Fig. 3).

It is experimentally proved that the existence of an optimal level of reducing (OLR) determines the minimal necessary number of input neurons of the perceptron.

The existence of an optimal level of structuring (OLS) is also proved where the difference (OLS-OLR) determines the extra number of the structuring operations for the input information already reduced as mentioned above.

Two heuristic algorithms are also suggested for a priori determination of the optimal levels, OLS and OLR, for a given set of graphics objects for recognition [9].

As a matter of further research it should be pointed the robustness of the proposed algorithms and the recognition rate comparison with other recognition approaches.

References

1. Progress in Handwriting Recognition. A. C. Downton and S. Impedovo, Eds., London, World Scientific, 1997.
2. Fundamentals in Handwriting Recognition. – In: S. Impedovo (ed.) NATO ASI Series, Series "F": Computer and System Sciences. Vol. 124. Springer Verlag, Berlin, 1994.
3. Duda, R., P. Hart. Pattern Classification and Scene Analysis. Wiley-Inter science Publication, 1973 (Russian edition: Moscow, Mir, 1976.).
4. Haykin, S. Neural Networks. Macmillan College Company, Inc., 1994.
5. Masters, T. Signal and Image Processing with Neural Networks. John Wiley & Sons, Inc., 1994.
6. Benedetto, J., M. Frazier. Wavelets – Mathematics and Applications. CRC Press, 1994, 1-20.
7. Daubechies, I. Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics. Philadelphia, PE, 1992.
8. Mallat, S. A theory for multiresolution signal decomposition: the wavelet representation. – IEEE Trans. on Pattern and Machine Intelligence, Vol. II, 1989, No 7.
9. Milanov, S., D. Dimov. Wavelets' Application for Structuring of Text Images. – In: Working papers series of IIT-BAS, ISSN-1310-652X, WP/69, Dec. 1998.

Приложение на вълнови трансформации при разпознаването на текст чрез невронни мрежи

Димо Димов, Сергей Миланов

Институт по информационни технологии, 1113 София, E-mail: dtdim@iinf.bas.bg

(Р е з ю м е)

Разглежда се използването на вълновите трансформации за предварителна обработка на графични изображения, подлежащи на разпознаване. За традиционния тип класификатор – нелинеен трислоен перцептрон, се изследва възможността за предварително структуриране и редуциране на информацията, подавана на входа му, с цел подобряване на разпознаващата способност. Анализът на експериментите показва, че могат да се дефинират оптимални нива на редуциране и структуриране на входната информация. Предлага се евристичен алгоритъм за априорна оценка на нивата, с който пряко се изчислява броят на входните неврони на перцептрона.