Computer Science. Vol. 3. 2001 004.032.26

Adrian Horzyk[*]

FAST AUTOMATIC CONFIGURATION OF ARTIFICIAL NEURAL NETWORKS USED FOR BINARY PATTERNS RECOGNITION

1. Introduction

Artificial neural networks can be used in many applications instead of the other sophisticated methods, but for a long time of learning and problems with finding proper architectures of neural networks makes it difficult to take advantages of neural computations. In order to reduce the time necessary for preparing responding neural network architecture to a specific recognition problem given by a learning sequence, some methods have to be developed. Architecture always reflects an ability of adapting and generalization of a neural network. In order to make meaningfully reductions of architecture so that the neural network furthermore solves a given problem of recognition is not easy. Moreover, while the reductions of architecture are done, a quality of them has to be appraised. We already know that reductions are possible, but this is difficult to determine a quality of them in general.

In this paper the limitation of qualitative reductions of neural network architecture is considered. There are a quality of recognition and a quality of generalization for described neural networks and for binary patterns recognition defined. The quality of recognition and the quality of generalization of such an automatically configured neural network can be automatically computed. These qualities and a number of synapses are always related. Weights of such a configured neural network are computed afterwards. In order to get an automatically configured neural network with automatically computed weights, two runs over a learning sequence are required. An input of a neural network can be a vector, a matrix, as well as any other structure, e.g. a log-polar structure. One limitation of this method is that it can be used only to identifying patterns of a learning sequence with qualitative generalization of noised inputs. Despite this limitation, the method can satisfy some groups of problems of recognition for its ability of determining a quality of recognition, a quality of generalization and a very fast method of automatic configuration of neural network architecture.

* Institute of Automatics, University of Mining and Metallurgy, Cracow

2. The estimation method

The proposed method estimates binary features of all given pattern first and then builds up the network basing on the computed estimations. In order to estimate binary features of patterns, the two matrices T and F have to be computed. The matrices T and F have the same dimensions (I, J) as the patterns of a given learning sequence. The matrix T consisting of some amount of true pixels and the matrix F consisting of some amount of false pixels, at the specified coordinates of all pattern matrices are defined as follows

$$\forall i,j \quad T[i,j] = \left\| p_k[i,j]\colon p_k[i,j] = true \,\&\, k = 1,\dots,N \right\|,$$

where N – number of patterns of a learning sequence,

$$\forall i,j \quad F[i,j] = \left\| p_k[i,j]\colon p_k[i,j] = false \,\&\, k = 1,\dots,N \right\|.$$

For the example of the 26-letter alphabet (shown partially in the Fig. 1), the two matrices T and F (shown in the Tab. 1) are computed.

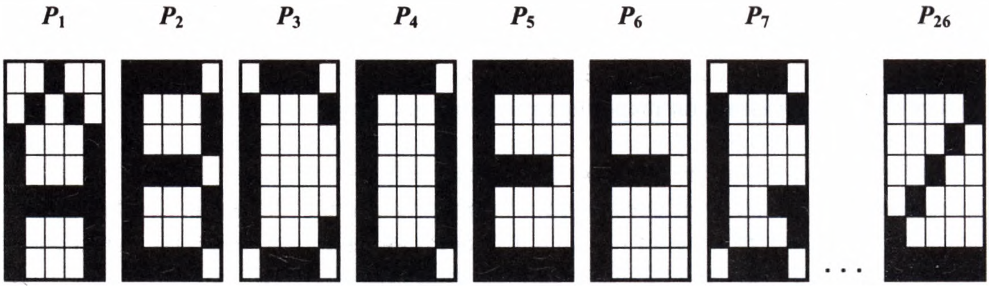

Fig. 1. The simple 26-letter alphabet: A, B, C, ..., Z (shown partially)

Table 1

The matrices T and F for the considered example of the simple 26-letter alphabet

$T =$

20	15	16	15	15
21	5	2	6	18
20	4	6	5	13
19	8	14	8	10
19	3	8	5	13
22	2	4	7	13
15	14	16	12	12

$F =$

6	11	10	11	11
5	21	24	20	8
6	22	20	21	13
7	18	12	18	16
7	23	18	21	13
4	24	22	19	13
11	12	10	14	14

Basing on these two matrices and on the number of patterns of a learning sequence (N), a binary feature estimator (E_k) of all pixels for each pattern can be computed as follows:

$$\forall k=1,\dots,N \quad \forall i,j \quad E_k[i,j]=\begin{cases}\dfrac{+1}{T[i,j]} & \text{if } P_k[i,j]=\text{true},\\[2ex] \dfrac{-1}{F[i,j]} & \text{if } P_k[i,j]=\text{false},\end{cases}$$

where

$$\forall k=1,\dots,N \quad \forall i,j \quad E_k[i,j]\in\left\langle -1,\frac{-1}{N}\right\rangle \text{ or } \left\langle \frac{1}{N},1\right\rangle .$$

The greater absolute $E_k[i, j]$ value indicates the greater importance of the pixel for a given pattern. For each pattern a group consisting of pixels with large absolute $E_k[i, j]$ values exists (e.g. the fields colored gray in the Tab. 2), which is the most representative for a given pattern in a view of a learning sequence. These groups are fundamental for the following computations.

Table 2

The matrix E_1 for the pattern „A"

$E_1 =$

–0,167	–0,091	0,063	–0,091	–0,091
–0,200	0,200	–0,042	0,167	–0,125
0,050	–0,045	–0,050	–0,048	0,077
0,053	–0,056	–0,083	–0,056	0,100
0,053	0,333	0,125	0,200	0,077
0,045	–0,042	–0,045	–0,053	0,077
0,067	–0,083	–0,100	–0,071	0,083

The main excitations

The main inhibitions

3. The quality of recognition and the quality of generalization

In order to obtain the best quality of recognition and the best quality of generalization no synaptic reductions are possible. To reduce a number of synapses, the quality of recognition and the quality of generalization have to be reduced as well. It is always a compromise between a number of synapses and these qualities factors. The quality of recognition (Q_R) and the quality of generalization (Q_G) can be defined in the following way:

$$Q_R=\frac{\min\limits_{k=1,\dots,N}\left(\min\limits_{p=1,\dots,N\,\&\,p\neq l}\left(\max\limits_{k=1,\dots,N}\left|Out_k^R[l]\right|-\left|Out_k^R[p]\right|\right)\right)}{\min\limits_{k=1,\dots,N}\left(\min\limits_{p=1,\dots,N\,\&\,p\neq l}\left(\max\limits_{k=1,\dots,N}\left|Out_k^F[l]\right|-\left|Out_k^F[p]\right|\right)\right)},$$

$$Q_G = \frac{\underset{k=1,\dots,N}{average}\left(\underset{p=1,\dots,N\&p\neq 1}{\min}\left(\underset{k=1,\dots,N}{\max}\left|Out_k^R[l]\right| - \left|Out_k^R[p]\right|\right)\right)}{\underset{k=1,\dots,N}{average}\left(\underset{p=1,\dots,N\&p\neq 1}{\min}\left(\underset{k=1,\dots,N}{\max}\left|Out_k^F[l]\right| - \left|Out_k^F[p]\right|\right)\right)},$$

where:

$Out_k^R[l]$ – the l-th output of the reduced neural network for the excitation of the k-th pattern of a learning sequence,

$Out_k^R[l]$ – is the l-th output of the full-connected neural network for the excitation of the k-th pattern of a learning sequence.

In order to obtain unequivocal recognition for all patterns of a learning sequence the quality of recognition Q_R have to be positive ($Q_R > 0$).

The problem can be formulated as follows: How to reduce as much as possible a number of synapses keeping the quality of recognition (Q_R) and the quality of generalization (Q_G) as high as possible? In order to satisfy both the requirements the three criteria of synapses reduction have been defined. While a synapsis with a low absolute $E_k[i, j]$ is reduced, the quality of recognition and the quality of generalization are not much declined, but when a synapsis with a high absolute $E_k[i, j]$ is reduced the quality of recognition and the quality of generalization are quickly lost. These circumstances are the basis for the reduction criteria defined as follows.

Reduction criteria

1) Criterion of Maximal Features (C_F)

   This criterion defines minimal estimation value $E_{\min}$, above which the synapses remain

$$E_{\min} = \frac{1}{1+(N-1)\cdot\frac{C_F}{100}}.$$

   This criterion reduces the number of synapses of individual patterns unequally. In some cases, this is strongly not recommended to employ only a few pixels of dominant $E_k[i, j]$ to recognize any pattern.

2) Criterion of the Minimal Number of Synapses (C_N)

   This criterion allows limiting a minimum number of synapses ($N_{\min}$), which always have to remain for each pattern of a learning sequence after reduction of synapses

$$N_{\min} = I \cdot J \cdot \frac{C_N}{100}.$$

3) Criterion of Minimal Precision of Recognition (C_P)

   In order to guarantee that some percentage of binary features ($P_{\min}$) of each pattern is stored in the synapses of a neural network, the two criteria formulated above are still insufficient. The third criterion is required and is defined as follows

$$P_{\min} = \sum_{i,j}\left|E_k[i,j]\right| \cdot \frac{C_p}{100}.$$

These three criteria may be used in different configurations. All of them are given in percentage. The three criteria formulated above define the *FNP (Feature, Number, Precision)* characteristic in the following way

$$FNP = C_F \mid C_N \mid C_P,$$

e.g. $FNP = 52|28|47$ means that: $C_F = 52\%$, $C_N = 28\%$, $C_P = 47\%$.

This *FNP* characteristic defines a neural network, which will be automatically configured for a given learning sequence. It also determines the quality of recognition Q_R and the quality of generalization Q_G, which will characterize the resultant neural network.

Usually, results of recognition are intuitively estimated and compared by using linear approaches. In order to recognize and compare the patterns in the intuitive scale a linear function of neuronal activity has to be used. Moreover, classification of negatives should reflect intuition. Both these intuitive requirements are satisfied by the use of the activity function f, which is defined as shown in the Figure 2.

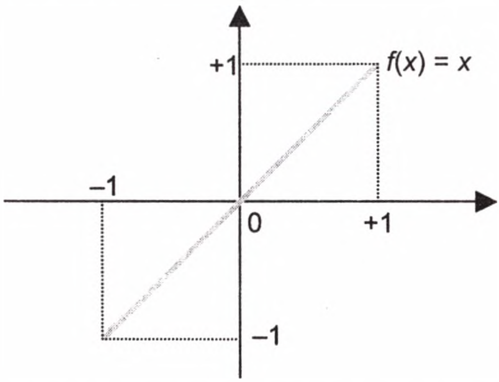

Fig. 2. The activity function of neurons takes the simple form: $f(x) = x$; $Dom\, f \in [-1,1]$, $Im\, f \in [-1,1]$

4. The neural network configuration

Automatic configuration of a neural network is now easy. After computing $E_k[i, j]$ for all pixels of all patterns of a given learning sequence only these synapses, which $E_k[i, j]$ satisfy *FNP* requirements remain and the rest of synapses is reduced (shown in the Fig. 3) by setting the appropriate $E_k[i, j]$ values to null (shown in the Tab. 3).

After synapses are created, the weights are computed from

$$W_k[i,j] = \frac{E_k[i,j]}{\sum_{m,n} \left|E_k[m,n]\right|}.$$

In result of such the computations, all outputs are normalized in $[-1, +1]$ interval (Tab. 4).

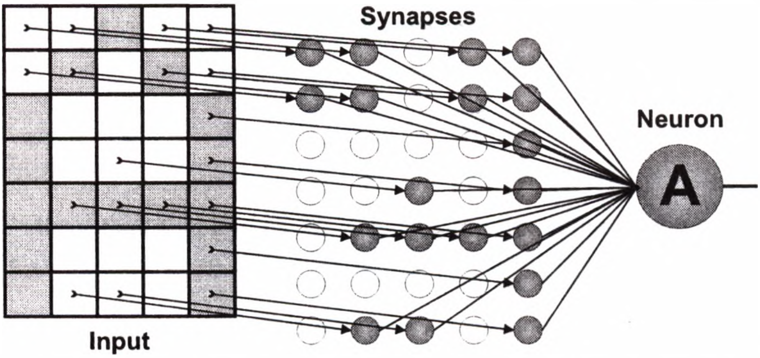

Fig. 3. The part of the neural network, which shows the reduced and left synapses for the learning pattern "A"

Table 3

The exemplary modified matrix E_1 for the learning pattern "A" after some reductions on the basis of the *FNP* characteristic

$E_1 =$

–0,167	–0,091	0	–0,091	–0,091
–0,200	0,200	0	0,167	–0,125
0	0	0	0	0,077
0	0	–0,083	0	0,100
0	0,333	0,125	0,200	0,077
0	0	0	0	0,077
0,067	–0,083	–0,100	0	0,083

Table 4

The resultant matrix of the weights for the learning pattern "A"

$E_1 =$

–0,068	–0,037		–0,037	–0,037
–0,081	0,081		0,068	–0,051
				0,031
		–0,034	0	0,040
	0,035	0,051	0,081	0,031
				0,030
	–0,034	–0,040		0,034

All these computations are done in two runs over a learning sequence, which makes this method very fast in comparison to the learning methods. The linearity of all computations makes them very simple, time saving and easy for intuitive interpretation.

5. Results of simulations

The computer implementation of this method has shown that for the example presented in this paper (i.e. the simple 26-letter alphabet: A, B, C, ..., Z) the quality of recognition $Q_R = 40.92\%$ is still positive ($Q_R > 0$) for the *FNP* = 52|28|47. The number of synapses has been reduced to 45% by the quality of generalization $Q_G = 48.20\%$.

When the quality of recognition is positive ($Q_R > 0$), there is a guarantee that all patterns of a given learning sequence are correctly and unequivocally recognized. The less correlation between the patterns of a learning sequence the greater reduction of synapses is possible and the less *FNP* values, which determine the resultant neural network, can be obtained. The quality of generalization Q_G is more important for recognition of noised patterns because it is a measure of ability to generalize. If the generalization of noised patterns is more important than the unequivocal recognition of patterns of a learning sequence and a greater reduction of synapses is necessary, then the reduction of synapses can pass the limit of $Q_R > 0$ (Tab. 5, 6). The presented method always remains the most important synapses after the *FNP* specification.

Table 5

The results of reduction for the 26-letter alphabet

FNP	Reduction of synapses	Q_G	Q_R
52|28|47	45%	48.20%	40.92%
45|20|40	34%	28.14%	0%
25|20|30	27%	11.86%	0%
15|15|15	22%	7.59%	0%
10|10|10	14%	5.68%	0%

Table 6

The results of reduction for the 30 bullets

FNP	Reduction of synapses	Q_G	Q_R
80|0|0	55%	71.89%	4.68%
58|21|0	32%	49.97%	0%
40|21|0	27%	29.92%	0%
25|14|25	19%	13.89%	0%

However, the patterns of a learning sequence have to be binary, the recognized pattern can be given in a scale of gray, i.e. between the values –1 (false) and 1 (true). Experiments have shown that such patterns are better recognized than their binarized form, e.g. the Figures 4, 5.

The patterns in negative are also recognized properly as the negatives of the learning patterns (e.g. Fig. 6).

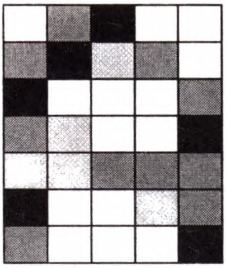

Fig. 4. The noised pattern “A”, given on the input of the neutral network in the gray scale, recognized as the learnining pattern “A” in 28.24%

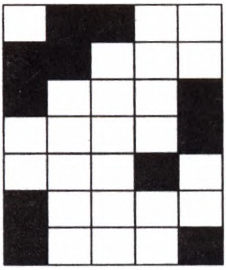

Fig. 5. The same noised pattern “A”, which was first binarized, given on the input of the neutral network, recognized as the learning pattern“A” in 13.4%

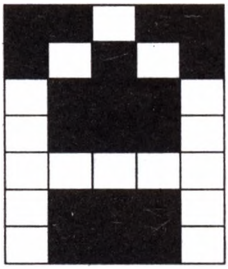

Fig. 6. The negative of the learning pattern “A”, given on the input of the neutral network recognized as “A” in “–100%”. The sign (–) designates that the pattern is more similar to the negative of the learning pattern “A” than to its positive. The value “100%” is a measure of its similarity to the learning pattern “A”

Some other simulations, e.g. the 30 bullets in raster 16 × 16, have shown similar ability of the presented method to reduce synapses (shown in the Tab. 6) as for the 26-letter alphabet considered in this paper.

The presented method for its ability to left the most important synapses can work on the small as well as huge learning sequences. Sometimes synaptic reduction is impossible, e.g. for the trivial learning sequences of logical functions: AND, OR, NOT, but for the most real learning sequences the reduction of synapses is possible.

6. Summary and conclusions

The powerful method of an automatically generated architecture of a neural network used to binary patterns recognition with an ability of synapses reduction in a way of minimally reducing a quality of generalization has been presented. This method has allowed reducing the number of synapses of the considered neural network architectures keeping the quality of recognition and the quality of generalization at high level. The presented method computes all weights in two runs over the learning sequence. The *FNP* characteristic, which unequivocally predefines a resultant structure of an automatically generated neural network for a given learning sequence has been defined. This method is not an antidote for all the problems of pattern recognition, but efficiently solves the class of problems presented in the paper.

Bibliography

[1] Hertz J., Krogh A., Palmer R.: *Introduction to the theory of neural computation.* Redwood City, Addison-Wesley Publishing Company, The Advanced Book Program 1991

[2] Horzyk A., Hybrid A.: *Neural Network Used for the Color Conversion from RGB to CMY and a New Method for Choosing a Learning Set.* Bydgoszcz, Image Processing & Communications, vol. 4, No. 1–2, ATR 1998

[3] Tadeusiewicz R.: *Sieci neuronowe.* Warszawa, Akademicka Oficyna Wydawnicza RM 1993

[4] Tadeusiewicz R., Flasiński M.: *Rozpoznawanie obrazów.* Warszawa, PWN 1991

[5] Żurada J., Barski M., Jędruch W.: *Sztuczne sieci neuronowe.* Warszawa, PWN 1996