American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-6, Issue-8, pp-345-359
www.ajer.org
Research Paper Open Access

Restoration of a Planar Image by Neural Network

Oleg Namicheishvili[1], Iveri Prangishvili, Jr.[2]
[1](*Department of Computer Engineering, Georgian Technical University, Georgia*)
[2](*Institute for Machine Learning at the Department of Computer Science, ETH Zurich, Switzerland*)
Corresponding Author: Oleg Namicheishvili

Abstract: *For the restoration problem with disturbances of a planar black-and-white video-image, we consider the restoring element with the aid of the formal neuron and propose the models of an optimal assignment of weights to its input dendrite channels. Also, the model of neuron adaptation is studied in detail using the stochastic approximation approach.*

Keywords: *Adaptation, dendrite weight, feedback, formal neuron, threshold*

Date of Submission: 19-08-2017 Date of acceptance: 31-08-2017

I. INTRODUCTION

This study deals with the topic that beyond any doubt presently attracts a great deal of interest. We consider the formal neuron whose inputs receive – via the dendrite channels $B_1, B_2, \ldots, B_n, B_{n+1}$ having various error probabilities q_i $\left(i=\overline{1,n+1}\right)$ – different binary versions $X_1, X_2, \ldots, X_n, X_{n+1}$ of one and the same random binary signal X . It is associated with a certain cell of the matrix that contains a video-image. It is assumed that $X=+1$ if a point of the video-image matches the matrix cell or $X=-1$ otherwise. The neuron is expected to restore the correct input signal X or, as we say, to make a decision Y on the basis of these $n+1$ dendrite versions $X_1, X_2, \ldots, X_n, X_{n+1}$. When the binary signal X is delivered to the inputs of the restoring element via the equally reliable channels, the decision making with regard for predominance of some value among other versions, i.e. according to the majority principle, was for the first time described by J. von Neumann [1].

When the input channels have different reliabilities, the restoration of a correct signal needs the adaptation of the formal neuron. The adaptation is interpreted as the process of control of weights a_i $\left(i=\overline{1,n+1}\right)$ of the neuron dendrite inputs so that these weights could match the current probabilities q_i $\left(i=\overline{1,n+1}\right)$ of errors of the dendrite channels. Any weight a_i is an arbitrary real number $\left(-\infty < a_i < +\infty\right)$. Such a control is aimed at making more reliable dendrites produce a greater influence on decision making (i.e. on the restoration of a correct signal) as compared with less reliable inputs.

Restoration (Fig. 1) is performed by weighted voting [2] by means of the relation

$$Y=\operatorname{sgn}\left(\sum_{i=1}^{n+1} a_i X_i\right)=\operatorname{sgn} Z$$

where

$$Z=\sum_{i=1}^{n+1} a_i X_i \ .$$

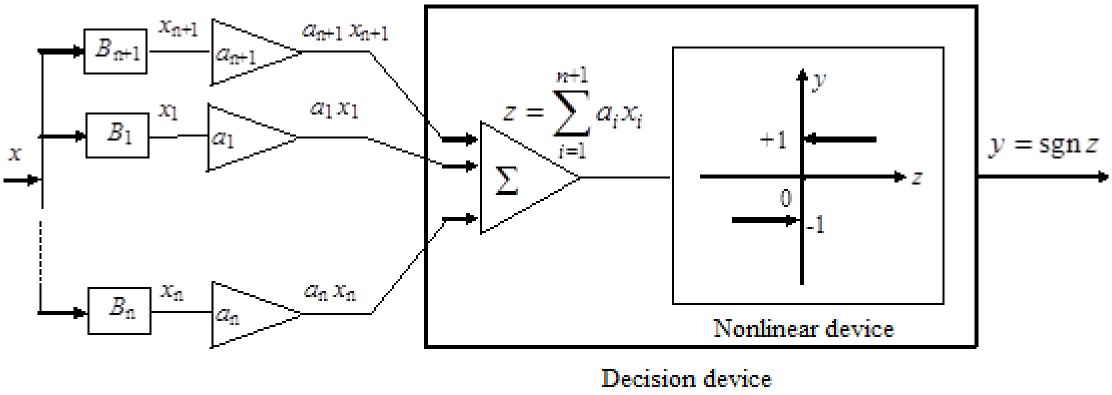

Fig. 1 The formal neuron in the role of a restoring device

II. SELECTION OF WEIGHTS FOR DENDRITE INPUT CHANNELS

Suppose there are two subsets (classes) of objects Ω_1 and Ω_2 which differ from each other according to some qualitative criterion. A priori probabilities $q_{n+1} = P(\Omega_1)$ and $1-q_{n+1} = P(\Omega_2)$ of the emergence of these two classes correspond to the supply of the artificial neuron with a random binary value X for recognition. Two possible values of X are coded as $+1$ and -1, respectively. Each object of the sesubsets is characterized by a set of qualitative criteria – a collection of $n+1$ number of binary versions of one and the same value X :

$$X_1, X_2, \ldots, X_n, X_{n+1}.$$

The nomenclature of these criteria is the same for Ω_1 and Ω_2, but the numerical values of parameters X_i may be different and may coincide for some individuals $B_i\left(i=\overline{1,n+1}\right)$. In the sequel, the realizations of binary random values X and $X_i\left(i=\overline{1,n+1}\right)$ are denoted by the respective small letters. Thus at the inputs («dendrites») of the artificial neuron we obtain the vector of observations

$$\vec{x} = \left(x_1, x_2, \ldots, x_n, x_{n+1}\right)^T$$

where T is the symbol of transposition of the row-vector into the column-vector. The vector of observations emerges according to the probabilities

$$\left.\begin{aligned} P(\vec{x}/\Omega_1) \equiv f_1(\vec{x}) &= \prod_{i=1}^{n+1} q_i^{\frac{1-x_i}{2}} \cdot (1-q_i)^{\frac{x_i+1}{2}} \\ P(\vec{x}/\Omega_2) \equiv f_2(\vec{x}) &= \prod_{i=1}^{n+1} q_i^{\frac{x_i+1}{2}} \cdot (1-q_i)^{\frac{1-x_i}{2}} \end{aligned}\right\}$$

where for the pseudorandom parameter X_{n+1} the realization x_{n+1} is invariably (-1) :

$$x_{n+1} \equiv -1.$$

Let us introduce the so-called discriminant function

$$Z = a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_{n+1}x_{n+1} = \sum_{i=1}^{n+1} a_1x_i$$

where $a_1\left(i=\overline{1,n+1}\right)$ are some real constants $\left(-\infty < a_1 < +\infty, i=\overline{1,n+1}\right)$ called *dendrite weights* of the artificial neuron, i.e. weights of objects $B_1\left(i=\overline{1,n+1}\right)$, and $\vec{x} = \left(x_1, x_2, \ldots, x_n, x_{n+1}\right)^T$ is the vector of observations which has to be attributed either to Ω_1 when $X=+1$ or to Ω_2 when $X=-1$. Thus, the multi-dimensional space is projected on to the one-dimensional one.

If the distribution were normal with respect to each parameter $X_i\left(i=\overline{1,n+1}\right)$, then the variable Z would also have a normal distribution. Assuming in rough approximation that this is so, thestandard classification procedure takes the following form of logical reasoning:

$$\left.\begin{aligned} a_1x_1+a_2x_2+\cdots+a_nx_n+a_{n+1}x_{n+1}=\sum_{i=1}^{n+1}a_1x_i\ge 0\Rightarrow x\in\Omega_1 \\ a_1x_1+a_2x_2+\cdots+a_nx_n+a_{n+1}x_{n+1}=\sum_{i=1}^{n+1}a_1x_i<0\Rightarrow x\in\Omega_2 \end{aligned}\right\}$$

which can also be equivalently written as

$$\left.\begin{aligned} & a_1x_1+a_2x_2+\cdots+a_nx_n=\sum_{i=1}^{n}a_1x_i\ge \Theta\Rightarrow x\in\Omega_1 \\ & a_1x_1+a_2x_2+\cdots+a_nx_n=\sum_{i=1}^{n}a_1x_i<\Theta\Rightarrow x\in\Omega_2 \\ & a_{n+1}\equiv\Theta \\ & x_{n+1}\equiv -1 \end{aligned}\right\}$$

where $\Theta\equiv a_{n+1}$ is some real constant $\left(-\infty<\Theta<+\infty\right)$ called the *threshold* [3] of the neuron.

III. GENERALIZED DISTANCE MAXIMUM CRITERION

Quite reasonably, there arises the question of finding weights $a_i\left(i=\overline{1,n+1}\right)$ for which the classification error is minimal. What does the erroneous classification mean?

If we consider the hypothetical case where the variable Z has the same normal distribution as the parameters $X_i\left(i=\overline{1,n+1}\right)$, then it turns out that for the classes Ω_1 and Ω_2 the density $f\left(z/\Omega_{1,2}\right)$ of the distribution of probabilities of a random variable Z has the form as shown in Fig.2.

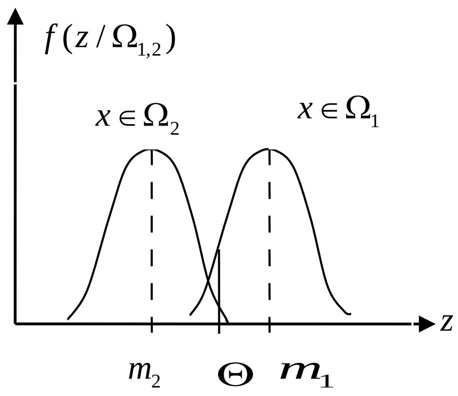

Fig. 2 Density $f\left(z/\Omega_{1,2}\right)$ of the distribution of probabilities of a random variable Z in the classes Ω_1 and Ω_2

The mathematical expectation for $Z=\sum_{i=1}^{n+1}a_i\cdot X_i$ in the class Ω_1, where $X=+1$, can be found by means of the mathematical expectation of a discrete random variable $X\cdot X_i$ that takes the value $x\cdot x_i=-1$ with probability q_i and the value $x\cdot x_i=+1$ with probability $1-q_i$. On the one hand, for $X=+1$ the mathematical expectation $M[Z]=M[X\cdot Z]$ and, on the other hand,

$$M\left[\sum_{i=1}^{n+1}a_i\cdot(X\cdot X_i)\right]=\sum_{i=1}^{n+1}a_i\cdot M[X\cdot X_i].$$

Therefore

$$m_1 \equiv M[Z/\Omega_1] = \sum_{i=1}^{n+1} a_i \cdot M[X \cdot X_i] = \sum_{i=1}^{n+1} a_i \cdot \left((-1)\cdot q_i + (+1)\cdot(1-q_i)\right) = \sum_{i=1}^{n+1} a_i \cdot (1-2q_i). \quad (1)$$

The mathematical expectation for $Z = \sum_{i=1}^{n+1} a_i \cdot X_i$ in the class Ω_2, where $X = -1$, can be expressed, as above, via the mathematical expectation of a discrete random value $X \cdot X_i$, that takes the value $x \cdot x_i = -1$ with probability q_i and the value $x \cdot x_i = +1$ with probability $1-q_i$. However, for $X = -1$ the mathematical expectation $M[Z] = -M[X \cdot Z]$ and, on the other hand, $M\left[\sum_{i=1}^{n+1} a_i \cdot (X \cdot X_i)\right] = \sum_{i=1}^{n+1} a_i \cdot M[X \cdot X_i]$. Therefore

$$m_2 \equiv M[Z/\Omega_2] = -\sum_{i=1}^{n+1} a_i \cdot M[X \cdot X_i] = -\sum_{i=1}^{n+1} a_i \cdot \left((-1)\cdot q_i + (+1)\cdot(1-q_i)\right) = \sum_{i=1}^{n+1} a_i \cdot (2q_i - 1). \quad (2)$$

From the comparison of relations (1) and (2) it follows that $m_1 = -m_2$ and therefore

$$m_1 - m_2 = 2 \cdot \sum_{i=1}^{n+1} a_i (1-2q_i). \quad (3)$$

The dispersion for a continuous random value $Z = \sum_{i=1}^{n+1} a_i \cdot X_i$ in the class Ω_1, where $X = +1$, can also be expressed via the dispersion of a discrete random variable $X \cdot X_i$ that takes the value $x \cdot x_i = -1$ with probability q_i and the value $x \cdot x_i = +1$ with probability $1-q_i$.

On the one hand, for $X = +1$ the dispersion $D[X \cdot Z] = (+1)^2 \cdot D[Z] = D[Z/\Omega_1] \equiv \sigma_{1Z}^2$ and, on the other hand, $D\left[\sum_{i=1}^{n+1} a_i \cdot (X \cdot X_i)\right] = \sum_{i=1}^{n+1} a_i^2 \cdot D[X \cdot X_i]$. Therefore for finding $D[X \cdot X_i]$ we should resort to Table 1, where the first row contains the realization $x \cdot x_i$ of a random value $X \cdot X_i$, and also its two values -1 and $+1$ that correspond to the recognition of a random binary variable X with error and without error. The second row for these two values contains the probabilities p_i of their realizations $p_{i1} = q_i$ and $p_{i2} = 1-q_i$, respectively. The third row contains for the discrete random value $\left(X \cdot X_i - M[X \cdot X_i]\right)$ its two possible values $(-1)-(1-2q_i) = -2+2q_i = -2(1-q_i)$ and $(+1)-(1-2q_i) = 2q_i$, since in any case $M[X \cdot X_i] = 1-2q_i$. The fourth row contains for the discrete random variable $\left(X \cdot X_i - M[X \cdot X_i]\right)^2$ its two possible values $\left(-2(1-q_i)\right)^2 = 4(1-q_i)^2$ and $(2q_i)^2 = 4q_i^2$. Finally, the fifth row contains the products of these two values of the discrete random variable $\left(X \cdot X_i - M[X \cdot X_i]\right)^2$ by the probabilities $p_i = q_i$ and $p_i = 1-q_i$ of the realizations of the values -1 and $+1$ of the discrete binary variable $X \cdot X_i$.

Table 1 Data for determining the dispersion (variation) of a random variable Z in the class Ω_1

	Values of a random variable $X \cdot X_i$	-1	$+1$
	Probabilities p_i of these values	q_i	$1-q_i$
	Mathematical expectation $M[X \cdot X_i]$	$(1-2q_i)$	
	Values of $\left(X \cdot X_i - M[X \cdot X_i]\right)$	$-1-(1-2q_i) = -2(1-q_i)$	$1-(1-2q_i) = 2q_i$
	Values of $\left(X \cdot X_i - M[X \cdot X_i]\right)^2$	$4(1-q_i)^2$	$4q_i^2$

	Values of $\left(X \cdot X_i - M[X \cdot X_i]\right)^2 \cdot p_i$	$4(1-q_i)^2 \cdot q_i$	$4q_i^2 \cdot (1-q_i)$

Summing the results of the last row, we obtain the dispersion $D\lfloor X \cdot X_i \rfloor$ of a random binary variable $X \cdot X_i$ in the class Ω_1, where $X = +1$:

$$D\lfloor X \cdot X_i \rfloor = 4(1-q_i)^2 \cdot q_i + 4q_i^2 \cdot (1-q_i) = 4q_i\left(1-q_i\right).$$

The substitution of $D\lfloor X \cdot X_i \rfloor$ into the formula

$$D[X \cdot Z] = (+1)^2 \cdot D[Z] = D[Z/\Omega_1] \equiv \sigma_{1Z}^2 = D\left[\sum_{i=1}^{n+1} a_i \cdot (X \cdot X_i)\right] = \sum_{i=1}^{n+1} a_i^2 \cdot D[X \cdot X_i]$$

yields

$$D[Z/\Omega_1] \equiv \sigma_{1Z}^2 = \sum_{i=1}^{n+1} 4a_i^2 q_i\left(1-q_i\right).$$

It is easy to guess that in the class Ω_2, where $X = -1$, we will have

$$D[X \cdot Z] = (-1)^2 \cdot D[Z] = D[Z/\Omega_2] \equiv \sigma_{2Z}^2 = D\left[\sum_{i=1}^{n+1} a_i \cdot (X \cdot X_i)\right] = \sum_{i=1}^{n+1} a_i^2 \cdot D[X \cdot X_i]$$

where the dispersion $D\lfloor X \cdot X_i \rfloor$ of a random binary variable $X \cdot X_i$ in the class Ω_2 for $X = -1$ will be expressed by the previous formula $D\lfloor X \cdot X_i \rfloor = 4q_i\left(1-q_i\right)$. Hence

$$D[Z/\Omega_2] \equiv \sigma_{2Z}^2 = \sum_{i=1}^{n+1} 4a_i^2 q_i\left(1-q_i\right).$$

Thus the dispersion σ_Z^2 of a random variable Z will be the same for the classes Ω_1 and Ω_2, since $D[Z/\Omega_1] \equiv \sigma_{1Z}^2 = D[Z/\Omega_2] \equiv \sigma_{2Z}^2$:

$$\sigma_{1Z}^2 = \sigma_{2Z}^2 \equiv \sigma_Z^2.$$

Finally,

$$\sigma_Z^2 = \sum_{i=1}^{n+1} 4a_i^2 q_i\left(1-q_i\right) = 4\sum_{i=1}^{n+1} a_i^2 q_i\left(1-q_i\right). \tag{3}$$

It is reasonable to choose dendrite weights a_i for which the mathematical expectations m_1 and m_2 would be as distant from each other as possible with respect to σ_Z^2 (i.e. it is required to maximally increase the distance between the humps having a minimal width). For this, it suffices to find a maximum of the value

$$\rho = \frac{\left(m_1 - m_2\right)^2}{\sigma_Z^2}. \tag{4}$$

From the condition $\rho = \max$ we determine the coefficients $a_1, a_2, \ldots, a_n, a_{n+1} \equiv \Theta$ of the discriminant function. The value ρ is called the *generalized distance* (or the *Mahalanobis distance* after the name of Indian statistician Mahalanobis who introduced [4] it in general usage) between two classes. It seems to be similar to the t-criterion by means of which the difference between two means is estimated, but this similarity is only outward.

In particular, if in expression (4) we use relations (1), (2) and (3), we obtain

$$\rho = \frac{\left(2 \cdot \sum_{i=1}^{n+1} a_i(1-2q_i)\right)^2}{4 \cdot \sum_{i=1}^{n+1} a_i^2 q_i (1-q_i)} = \frac{\left(\sum_{i=1}^{n+1} a_i(1-2q_i)\right)^2}{\sum_{i=1}^{n+1} a_i^2 q_i (1-q_i)}. \tag{5}$$

By considering ρ as a ratio of u to v, we find the derivative of this relation by the formula

$$\left(\frac{u}{v}\right)' = \frac{v \cdot u' - u \cdot v'}{v^2}, v \neq 0 .$$

As a result, we have

$$\frac{\partial \rho}{\partial a_i} = \frac{\left(\sum_{i=1}^{n+1} a_i^2 q_i(1-q_i)\right) \cdot \left(\left(2\sum_{i=1}^{n+1} a_i(1-2q_i)\right) \cdot \sum_{i=1}^{n+1}(1-2q_i)\right) - \left(\sum_{i=1}^{n+1} a_i(1-2q_i)\right)^2 \cdot \left(2 \cdot \sum_{i=1}^{n+1} a_i q_i(1-q_i)\right)}{\left(\sum_{i=1}^{n+1} a_i^2 q_i(1-q_i)\right)^2}$$

.

Optimal dendrite weights a_{im} $(i = \overline{1, n+1})$ with respect to a maximum of the Mahalanobis distance can be defined from the system of equations

$$\left.\begin{array}{l} \dfrac{\partial \rho}{\partial a_i} = 0 \\ i = \overline{1, n+1} \end{array}\right\},$$

which with the use of the preceding expression (provided that the denominator is not equal to zero) gives

$$\left.\begin{array}{l} \left(\sum_{i=1}^{n+1} a_i^2 q_i(1-q_i)\right) \cdot \left(\sum_{i=1}^{n+1}(1-2q_i)\right) - \left(\sum_{i=1}^{n+1} a_i(1-2q_i)\right) \cdot \left(\sum_{i=1}^{n+1} a_i q_i(1-q_i)\right) = 0 \\ i = \overline{1, n+1} \end{array}\right\}.$$

This relation can be rewritten in the form

$$\left.\begin{array}{l} \dfrac{\sum_{i=1}^{n+1}(1-2q_i)}{\sum_{i=1}^{n+1} a_i q_i(1-q_i)} = \dfrac{\sum_{i=1}^{n+1} a_i(1-2q_i)}{\sum_{i=1}^{n+1} a_i^2 q_i(1-q_i)} \\ i = \overline{1, n+1} \end{array}\right\}. \qquad (6)$$

This requirement *does not contradict* the assumption that

$$\left.\begin{array}{l} (1-2q_i) = a_i \cdot q_i(1-q_i) \\ i = \overline{1, n+1} \end{array}\right\}, \qquad (7)$$

since then there also holds the following statement

$$\left.\begin{array}{l} (1-2q_i) = a_i \cdot q_i(1-q_i) \\ a_i^2 \cdot q_i(1-q_i) = a_i \cdot (1-2q_i) \\ i = \overline{1, n+1} \end{array}\right\}.$$

Here the summation of the left- and right-hand parts gives

$$\left.\begin{array}{l} \sum_{i=1}^{n+1}(1-2q_i) = \sum_{i=1}^{n+1} a_i \cdot q_i(1-q_i) \\ \sum_{i=1}^{n+1} a_i^2 \cdot q_i(1-q_i) = \sum_{i=1}^{n+1} a_i \cdot (1-2q_i) \\ i = \overline{1, n+1} \end{array}\right\}$$

or, which is the same,

$$\left.\begin{aligned} &\sum_{i=1}^{n+1}(1-2q_i)=\sum_{i=1}^{n+1}a_i\cdot q_i(1-q_i)\\ &\sum_{i=1}^{n+1}a_i\cdot(1-2q_i)=\sum_{i=1}^{n+1}a_i^2\cdot q_i(1-q_i)\\ &i=\overline{1,n+1}\end{aligned}\right\}.$$

Termwise division of these two equalities gives relation (6):

$$\left.\begin{aligned} &\frac{\sum_{i=1}^{n+1}(1-2q_i)}{\sum_{i=1}^{n+1}a_i\cdot(1-2q_i)}=\frac{\sum_{i=1}^{n+1}a_i\cdot q_i(1-q_i)}{\sum_{i=1}^{n+1}a_i^2\cdot q_i(1-q_i)}\\ &i=\overline{1,n+1}\end{aligned}\right\}.$$

Thus the statement that assumption (7) does not contradict requirement (6) is proved.

But assumption (7) implies that the weights $a_i=a_{im}$ $(i=\overline{1,n+1})$ supplying a maximum $\rho_{\max}$ to the Mahalanobis distance ρ must be defined by the following relations

$$\boxed{\left.\begin{aligned} &a_{im}=\frac{1-2q_i}{q_i(1-q_i)}\\ &i=\overline{1,n+1}\end{aligned}\right\}.} \tag{8}$$

For such weights, by virtue of formula (5) the maximal value $\rho_{\max}$ of the Mahalanobis distance is

$$\rho_{\max}=\frac{\left(\sum_{i=1}^{n+1}a_{im}(1-2q_i)\right)^2}{\sum_{i=1}^{n+1}a_{im}^2q_i(1-q_i)}=\frac{\left(\sum_{i=1}^{n+1}\frac{1-2q_i}{q_i(1-q_i)}(1-2q_i)\right)^2}{\sum_{i=1}^{n+1}\left(\left(\frac{1-2q_i}{q_i(1-q_i)}\right)^2\cdot q_i(1-q_i)\right)}=\frac{\left(\sum_{i=1}^{n+1}\frac{(1-2q_i)^2}{q_i(1-q_i)}\right)^2}{\sum_{i=1}^{n+1}\frac{(1-2q_i)^2}{q_i(1-q_i)}}$$

which, upon reducing the numerator and the denominator, gives the following final result

$$\boxed{\rho_{\max}=\sum_{i=1}^{n+1}\frac{(1-2q_i)^2}{q_i(1-q_i)}.} \tag{9}$$

It is easy to verify that this value is half the difference (more exactly, half the absolute value of the difference) of the mathematical expectations m_1 and m_2 of a random sum Z for the classes Ω_1 and Ω_2 if the dendrite weights a_i $(i=\overline{1,n+1})$ of the neuron are chosen by relations (8):

$$\rho_{\max}=\frac{1}{2}|m_1-m_2|,\ \text{if } a_i=a_{im},\ i=\overline{1,n+1}.$$

IV. ENTROPY SENSITIVITY CRITERION

For an alternative approach, $n+1$ dendrites $B_1, B_2, \ldots, B_n, B_{n+1}$ can be interpreted as some source of binary information with entropy E defined by the formula

$$E=\sum_{i=1}^{n+1}E_i \tag{10}$$

where E_i $\left(i=\overline{1,n+1}\right)$ is the entropy of a discrete random variable $X \cdot X_i$ whose distribution is given in the form of a set of probabilities q_i and $\left(1-q_i\right)$, corresponding to its realizations $x \cdot x_i = -1$ and $x \cdot x_i = +1$, respectively.

Thus

$$E_i = -\left(q_i \cdot \ln q_i + \left(1-q_i\right) \cdot \ln\left(1-q_i\right)\right). \tag{11}$$

Hence

$$E = -\sum_{i=1}^{n+1}\left(q_i \cdot \ln q_i + \left(1-q_i\right) \cdot \ln\left(1-q_i\right)\right). \tag{12}$$

The weight a_i of the dendrite B_i $\left(i=\overline{1,n+1}\right)$ will serve as a measure of variation ∂E of the entropy E throughout the entire set depending on an increment ∂q_i of the probability q_i of the error of this concrete dendrite

$$\left.\begin{array}{l} a_i = \dfrac{\partial E}{\partial q_i} \\ i=\overline{1,n+1} \end{array}\right\}. \tag{13}$$

Applying expression (12) for E, by the entropy sensitivity criterion (definition) (13) the weight of the dendrite a_{ie} is written in the form

$$\boxed{\left.\begin{array}{l} a_{ie} = \ln \dfrac{1-q_i}{q_i} \\ i=\overline{1,n+1} \end{array}\right\}.} \tag{14}$$

V. RELATIONSHIP OF DENDRITE WEIGHTS DEFINED BY VARIOUS CRITERIA

To establish the relationship between a_{im} and a_{ie}, we write expression (8) in a somewhat different form

$$\left.\begin{array}{l} a_{im} = \dfrac{1-2q_i}{q_i(1-q_i)} = \dfrac{1-q_i}{q_i} - \dfrac{q_i}{1-q_i} \\ i=\overline{1,n+1} \end{array}\right\}.$$

Here, formally,

$$\left.\begin{array}{l} \dfrac{1-q_i}{q_i} = \exp\left(\ln \dfrac{1-q_i}{q_i}\right) = \exp\left(a_{ie}\right) \\ \dfrac{q_i}{1-q_i} = \exp\left(\ln \dfrac{q_i}{1-q_i}\right) = \exp\left(-\ln \dfrac{1-q_i}{q_i}\right) = \exp\left(-a_{ie}\right) \end{array}\right\}.$$

Using this fact in the preceding formula, we obtain

$$\left.\begin{array}{l} a_{im} = \exp(a_{ie}) - \exp(-a_{ie}) = 2\dfrac{\exp(a_{ie}) - \exp(-a_{ie})}{2} \\ i=\overline{1,n+1} \end{array}\right\}.$$

Here, based on the definition of a hyperbolic sine, we may have the representation

$$\left.\begin{array}{l} \dfrac{\exp(a_{ie}) - \exp(-a_{ie})}{2} = \sinh(a_{ie}) \\ i=\overline{1,n+1} \end{array}\right\}. \tag{15}$$

Finally, we obtain

$$\left.\begin{array}{l} a_{im} = 2\sinh(a_{ie}) \\ i = \overline{1, n+1} \end{array}\right\}. \qquad (16)$$

VI. GRAPHIC INTERPRETATION OF THE RESULTS COMPUTED FOR DENDRITE WEIGHTS

Figure 3 shows for the sake of comparison the graphs of relationship between (8) (i.e. (16)) and (14).

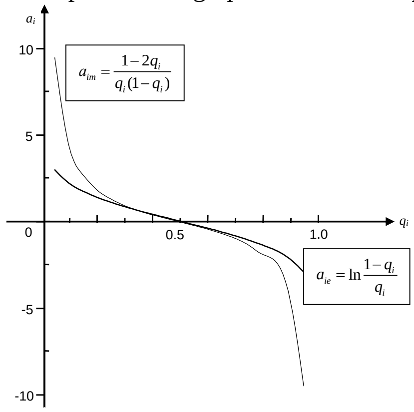

Fig. 3 Dependence of the weight a_i on the probability q_i for two optimality criteria

.

VII. STOCHASTIC APPROXIMATION FOR THE CONTROL OF DENDRITE WEIGHTS OF THE NEURON

Figure 4 presents the block-diagram of realization of the methods of continuous adaptation [5](both with and without feedback) by an algorithm of stochastic approximation type [6, 7] with the aim of computing an increment of the input weights.

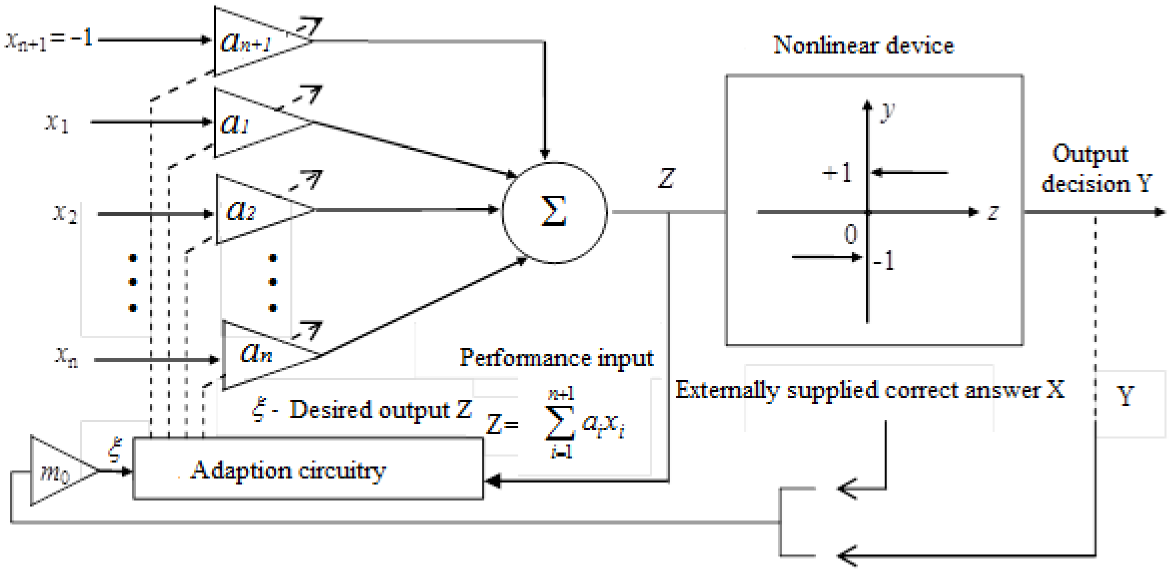

Fig. 4 The continuous adaptation block-diagram for computation of optimal dendrite weights

For a random increment $\Delta a_i(k) = a_i(k+1) - a_i(k)$ of the weight of the B_i dendrite $a_i(k)$ obtained at the iteration step $(k+1)$ we write

$$\left.\begin{array}{l}\Delta a_i(k)=\gamma_k\cdot X_i(k)\left(m_0X-\sum_{j=1}^{n+1}a_j(k)\cdot X_j(k)\right)\\ i=\overline{1,n+1}\end{array}\right\}. \quad (17)$$

Here γ_k means the fraction of errors, which are to be eliminated at each iteration step $(k+1)$. Because the dendrite weights keep being continuously refined, this fraction is supposed to decrease by the rule

$$\left.\begin{array}{l}\gamma_k=\dfrac{1}{k}\\ k=1,2,\ldots\end{array}\right\}. \quad (18)$$

γ_k being an element of the so-called harmonic sequence will satisfy the following conditions

$$\left.\begin{array}{l}\lim\limits_{k\to\infty}\gamma_k=0\\ \sum\limits_{k=1}^{\infty}\gamma_k=\infty\\ \sum\limits_{k=1}^{\infty}\gamma_k^2=\dfrac{1}{6}\pi^2\approx1.6449<\infty\end{array}\right\}. \quad (19)$$

The last relation in this system is the solution found by Leonard Euler as far back as 1735 for the so-called «Basel problem» to which the attention of European mathematicians was for the first time drawn by Basel professor of mathematics Jacob Bernoulli. Thus the problem reduced to finding a mathematical expectation of a random increment variable of increment (17) and deriving expressions for those ***weights*** which match the stationary state with zero mathematical expectations. As is well known, the convergence of the iteration process to this state is provided by the fulfillment of conditions (19).

The quantizing element (adaptation circuit) receives, on the one hand, the value $z=\sum_{1=1}^{n+1}a_ix_i$ of a random sum Z and, on the other hand, the value $\xi=m_0X$ (in the absence of feedback) or the value $\xi=m_0Y$ (in the presence of feedback), where m_0- is a given restriction imposed on the absolute value of the sum Z. The increment value of the dendrite weight is defined by the deviation of the desired m_0X (or m_0Y) from the real value of the sum Z. What is the advantage that makes this approach attractive?

There may occur a case where one dendrite suppresses the rest and completely determines the output signal. If the domain of weight values were not restricted, then the deciding element could be «***captured***» by one dendrite input which had no non-coincidences with the output solution during the adaptation process. If the new weight of the vote of the dominating dendrite exceeds the sum of weight values of votes of all other dendrites, then in the next adaptation act the signal of the dominating dendrite will coincide with the output signal. In case this situation turns out to persist, the dominating dendrite will determine the solution in many or even in all subsequent adaptation cycles. This will the «***usurpation***» of the right to make a decision, which is especially dangerous since at some moment of time the dominating dendrite may turn out to be unreliable [8, 9].

Let us transform equality (17) and write it in the following form

$$\left.\begin{array}{l}\Delta a_i(k)\}=\gamma_k\cdot m_0\cdot X\cdot X_i(k)-\gamma_k\cdot X_i(k)\cdot\sum\limits_{j=1}^{n+1}a_j(k)\cdot X_j(k)\\ i=\overline{1,n+1}\end{array}\right\}.$$

If here in the sum $\sum_{j=1}^{n+1}a_j(k)\cdot X_j(k)$ we pick out the term $a_i(k)X_i(k)$, then for the remaining part of the initial sum we can introduce the conditional notation

$$\sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)\cdot X_j(k).$$

Hence in the preceding relation there appears the three-component equality

$$\left.\begin{array}{l}\Delta a_i(k)\} = \gamma_k\cdot m_0\cdot X\cdot X_i(k) - \gamma_k\cdot a_i(k)\cdot X_i^2(k) - \gamma_k\cdot X_i(k)\cdot \sum\limits_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)\cdot X_j(k)\\ i=\overline{1,n+1}\end{array}\right\}. \tag{20}$$

If the errors of neuron dendrites are independent, then the mathematical expectation $M[\Delta a_i(k)]$ of the increment $\Delta a_i(k)$ is found as the sum of mathematical expectations of three separate summands of this increment.

In particular, for the first two mathematical expectations (a discrete random value and a constant) we have by definition

$$\left.\begin{array}{l} M\left[\gamma_k m_0\cdot XX_i(k)\right] = m_0\gamma_k\left[(-1)q_i + (+1)(1-q_i)\right] = m_0\gamma_k(1-2q_i)\\ M\left[\gamma_k a_i(k)\cdot X_i^2(k)\right] = \gamma_k a_i(k)\cdot M\left[X_i^2(k)\right] = \gamma_k a_i(k)\cdot M[1] = \gamma_k a_i(k)\end{array}\right\}. \tag{21}$$

To find the mathematical expectation of the summand $\gamma_k\cdot X_i(k)\cdot \sum\limits_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)\cdot X_j(k)$, we need to compute it separately for the classes Ω_1 and Ω_2, where the random variable X has respectively the values $X=+1$ and $X=-1$. In particular, for $X=+1$ the value $X_i=+1$ is realized with probability $(1-q_i)$ of the absence of an error, while the value $X_i=-1$ appears with probability q_i of an error, i.e.

$$M[X_i / X=+1] = (+1)\cdot(1-q_i) + (-1)\cdot q_i = 1-2q_i.$$

Analogously, for $X=-1$ the value $X_i=+1$ is realized with probability q_i of an error, while the value $X_i=-1$ appears with probability $(1-q_i)$ of the error absence, i.e.

$$M[X_i / X=-1] = (+1)\cdot q_i + (-1)\cdot(1-q_i) = 2q_i - 1.$$

Thus we have

$$M\left[\gamma_k\cdot X_i(k)\cdot \sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)\cdot X_j(k) \Big/ X=+1\right] = \gamma_k\cdot(1-2q_i)\cdot \sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)\cdot(1-2q_j),$$

$$M\left[\gamma_k\cdot X_i(k)\cdot \sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)\cdot X_j(k) \Big/ X=-1\right] = \gamma_k\cdot(2q_i-1)\cdot \sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)\cdot(2q_j-1).$$

In the latter case the result remains as before since

$$\gamma_k\cdot(2q_i-1)\cdot \sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)\cdot(2q_j-1) = \gamma_k\cdot(1-2q_i)\cdot \sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)\cdot(1-2q_j)\cdot$$

Thus, independently of the class Ω_1 or Ω_2, the mathematical expectation of the third summand in relation (20) has the form

$$M\left[\gamma_k\cdot X_i(k)\cdot \sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)\cdot X_j(k)\right] = \gamma_k\cdot(1-2q_i)\cdot \sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)\cdot(1-2q_j). \tag{22}$$

Taking results (21) and (22) into account and using relation (20), for the mathematical expectation $M[\Delta a_i(k)]$ of a weight increment we will have

$$\left.\begin{array}{l} M[\Delta a_i(k)] = \gamma_k m_0 \cdot (1-2q_i) - \gamma_k a_i(k) - \gamma_k (1-2q_i) \cdot \sum_{\substack{j=1 \\ j\neq i}}^{n+1} a_j(k)\cdot(1-2q_j) \\ i=\overline{1,n+1} \end{array}\right\}.$$

This relation can be rewritten in the following form, too,

$$\left.\begin{array}{l} M[\Delta a_i(k)] = \gamma_k \left\{ (1-2q_i)\left[m_0 - \sum_{\substack{j=1 \\ j\neq i}}^{n+1} a_j(k)\cdot(1-2q_j) \right] - a_i(k) \right\} \\ i=\overline{1,n+1} \end{array}\right\}.$$

Since here

$$\sum_{\substack{j=1 \\ j\neq i}}^{n+1} a_j(k)\cdot(1-2q_j) = \left(\sum_{j=1}^{n+1} a_j(k)\cdot(1-2q_j) \right) - a_i(k)\cdot(1-2q_i),$$

the preceding relation transforms to the form

$$\left.\begin{array}{l} M[\Delta a_i(k)] = \gamma_k \left\{ (1-2q_i)\left[m_0 - \sum_{j=1}^{n+1} a_j(k)\cdot(1-2q_j) \right] + a_i(k)\cdot(1-2q_i)^2 - a_i(k) \right\} \\ i=\overline{1,n+1} \end{array}\right\}$$

where

$$a_i(k)\cdot(1-2q_i)^2 - a_i(k) = -4a_i(k)q_i(1-q_i).$$

Finally, we have

$$\left.\begin{array}{l} M[\Delta a_i(k)] = \gamma_k \left\{ (1-2q_i)\left[m_0 - \sum_{j=1}^{n+1} a_j(k)\cdot(1-2q_j) \right] - 4a_i(k)q_i(1-q_i) \right\} \\ i=\overline{1,n+1} \end{array}\right\}.$$

For the state of equilibrium in which the mathematical expectations of weight increments take zero values we obtain

$$\left.\begin{array}{l} (1-2q_i)\left[m_0 - \sum_{j=1}^{n+1} a_j(k)\cdot(1-2q_j) \right] - 4a_i(k)q_i(1-q_i) = 0 \\ i=\overline{1,n+1} \end{array}\right\}.$$

Denoting the weights for which this state occurs by the symbol $\hat{a}_i$ $\left(i=\overline{1,n+1}\right)$, from the last equations we obtain

$$\left.\begin{array}{l} \hat{a}_i = \dfrac{1-2q_i}{q_i(1-q_i)} \cdot \dfrac{m_0 - \sum_{j=1}^{n+1} \hat{a}_j(1-2q_j)}{4} \\ i=\overline{1,n+1} \end{array}\right\}.$$

In the structure of this formula, the first co-factor (8) represents the weights a_{im}, which supply maximal value (9) to the Mahalanobis distance between the sets of values of a random sum $Z=\sum_{i=1}^{n+1} a_i X_i$ in the classes Ω_1 and Ω_2. By the optimal weights a_{ie} defined by the entropy sensitivity formula (14) they are related via (16): $a_{im} = 2\sinh(a_{ie})$.

As to the second co-factor, it does not depend on the index i and is completely defined by the weights, which match the state of equilibrium.
Thus we finally have

$$\left.\begin{array}{l}\hat{a}_i = \gamma \cdot a_{im} = \gamma \cdot 2\sinh(a_{ie}) \\ i = \overline{1, n+1}\end{array}\right\} \quad (23)$$

where, according to the notation,

$$\gamma = \frac{m_0 - \sum_{j=1}^{n+1} \hat{a}_j (1-2q_j)}{4}. \quad (24)$$

Using sequentially the expression $\hat{a}_i = \gamma \cdot a_{im}$ in notation (24), and formulas (8) and (9) we have

$$\gamma = \frac{m_0 - \sum_{j=1}^{n+1} \hat{a}_j (1-2q_j)}{4} = \frac{m_0 - \gamma \cdot \sum_{j=1}^{n+1} a_{jm} (1-2q_j)}{4} =$$

$$= \frac{m_0 - \gamma \cdot \sum_{j=1}^{n+1} \frac{1-2q_j}{q_j(1-q_j)}(1-2q_j)}{4} = \frac{m_0 - \gamma \cdot \sum_{j=1}^{n+1} \frac{(1-2q_j)^2}{q_j(1-q_j)}}{4} = \frac{m_0 - \gamma \cdot \rho_{\max}}{4}.$$

The result

$$\gamma = \frac{m_0 - \gamma \cdot \rho_{\max}}{4}$$

implies

$$\gamma = \frac{m_0}{4 + \rho_{\max}} \quad (25)$$

where, as has been mentioned above, m_0 is a given restriction on the absolute value of the sum Z.

Thus, when using the method described in this paper, the weights of dendrite channels in the formal neuron which match the state of equilibrium are proportional to the weights computed by the criterion of maximum of a generalized distance or, which is the same, by the hyperbolic sine of weights which estimate entropy sensitivity. Simultaneously, these weights are inversely proportional to the maximum $\rho_{\max}$ of the Mahalanobis distance ρ between the distributions of a random sum Z in the classes Ω_1 (when $X = +1$) and Ω_2 (when $X = -1$).

The argumentation is also well-grounded for adaptation with feedback, but in that case instead of the dendrite error probability one has to apply the probability of deviation of the dendrite from the global solution of the whole neuron.

VIII. CONCLUSION

A. The first result

The weights of the neuron dendrite inputs a_{ie}, calculated by the criterion of entropy sensitivity, are given by the following formula

$$\left.\begin{array}{l}a_{ie} = \ln\frac{1-q_i}{q_i} \\ i = \overline{1, n+1}\end{array}\right\}.$$

B. The second result

The weights of the neuron dendrite inputs a_{im}, calculated by the criterion of the maximum Mahalanobis distance, are given by the following formula

$$\left.\begin{array}{l} a_{im} = \dfrac{1-2q_i}{q_i(1-q_i)} \\ i = \overline{1, n+1} \end{array}\right\}.$$

Recall that the Mahalanobis distance is given by the relationship

$$\rho = \frac{(m_1 - m_2)^2}{\sigma_z^2}$$

where m_1 and m_2 are the mathematical expectations of the random sum Z at $X = +1$ and $X = -1$ respectively

$$\left.\begin{array}{l} m_1 = M\left[Z / X = +1\right] \\ m_2 = M\left[Z / X = -1\right] \end{array}\right\}.$$

As for the value of the nominator of the Mahalanobis distance, it represents the dispersion of the same sum. The dispersion is identical at both conditions, that is

$$D\left[Z / X = +1\right] = D\left[Z / X = -1\right] \equiv \sigma_Z^2.$$

C. The third result

The link between the above-mentioned weights is established on the basis of monotonic rearrangement by a hyperbolic sine

$$\left.\begin{array}{l} a_{im} = 2\sinh(a_{ie}) \\ i = \overline{1, n+1} \end{array}\right\}.$$

D. The fourth result

The maximum value of the Mahalanobis distance is determined by the following expression

$$\rho_{\max} = \sum_{i=1}^{n+1} \frac{(1-2q_i)^2}{q_i(1-q_i)}.$$

E. The fifth result

In the process of continuous adaptation, when the increment of weights is formed by the stochastic approximation algorithm, the weights that are proportional to those that give the maximum to the Mahalanobis distance are determined

$$\left.\begin{array}{l} \hat{a}_i = \gamma \cdot a_{im} = \gamma \cdot 2\sinh(a_{ie}) \\ \gamma = \dfrac{m_0}{4 + \rho_{\max}} \\ i = \overline{1, n+1} \end{array}\right\}$$

where m_0 is a given restriction on the absolute value of the sum Z .

REFERENCES

[1] J. von Neumann, Probabilistic logics and the synthesis of reliable organisms from unreliable components, in C. Shannon and J. McCarthy (Editors), Automata Studies (Princeton University Press,1956). 43-98.

[2] Archil Prangishvili, Oleg Namicheishvili, and Archil Elizbarashvili**,** Neural Networks (Tbilisi: Technical University Publishing House, 2007, in Georgian).

[3] M.L. Dertouzos, Threshold Logic: a Synthesis Approach (Cambridge: the MIT Press, 1965).

[4] P.C. Mahalanobis, On the Generalized Distance in Statistics, Proceedings of the National Institute of Sciences of India, 12, 1936,. 49-55.

[5] W.H. Pirce, Failure-Tolerant Computer Design (New York and London: Academic Press, 1965). See Appendix 2, pp. 201-203, and Chapter IV, pp. 56-58.

[6] H. Robbins and S. Monro, A Stochastic Approximation Method, The Annals of Mathematical Statistics, 22(3), 1951, 400-407.

[7] B. Widrow and M.E. Hoff, Jr., Adaptive Switching Circuits, IRE WESCON Convention Record, New York: IRE, Part 4, 1960, 96-104.

[8] A.I. Prangishvili, O.M. Namicheishvili, Formal Neuron and the Models of its Adaptation in the Redundancy Theory, Information and Computer Technologies: Theory and Practice: Proceedings of the International Scientific Conference Devoted to the 80th

Anniversary of Academician I.V. Prangishvili, Series: Computer Science, Technology and Applications, New York: Nova Science Publishers 2012, 1-9.

[9] A.I. Prangishvili, O.M. Namicheishvili, M.A. Gogiashvili, Two Methods of Obtaining a Minimal Upper Estimate for the Error Probability of the Restoring Formal Neuron, Proceedings of the International Conference on Mathematical Methods, Mathematical Models and Simulation in Science and Engineering MMSSE 2014, Series 23 - Mathematics and Computers in Science and Engineering, Interlaken, Switzerland, 2014, 37-41.

Oleg Namicheishvili. "Restoration of a Planar Image by Neural Network." American Journal of Engineering Research (AJER), vol. 6, no. 8, 2017, pp. 345–359.