UDC 004.032.26::004.93+519.8

OPTIMAL HIDDEN LAYER NEURONS NUMBER IN TWO-LAYER PERCEPTRON AND PIXEL-TO-TURN STANDARD DEVIATIONS RATIO FOR ITS TRAINING ON PIXEL-DISTORTED TURNED 60×80 IMAGES IN TURNED OBJECTS CLASSIFICATION PROBLEM

V. V. Romanuke

Khmelnytskyy National University

vul. Instytutska, 11, Khmelnytskyy, 29000, Ukraine. E-mail: romanukevadimv@mail.ru

Purpose. To optimize the two-layer perceptron by its two parameters for turned objects classification problem. The first parameter is the number of hidden layer neurons. The second one is ratio of pixel-to-turn standard deviations in forming the training set. The general totality is of plane objects imaged monochrome, and thus the two-layer perceptron is trained to classify turned 60×80 monochrome images. **Results.** The optimal hidden layer neurons number in the two-layer perceptron has been found along with pixel-to-turn standard deviations ratio for training on pixel-distorted turned 60×80 images. The pair of those optimal parameters, however, is not unique, and other versions of optimal pairs are listed. Generally, for classifying turned objects, the optimal hidden layer neurons number shall be adjusted to the set of 131 elements starting at 220 and ending at 350. The best pixel-to-turn standard deviations ratio is a value from the segment $[0.03; 0.1]$, although roughly. For 60×80 images, the most appropriate ratio is a value from the segment $[0.04; 0.06]$. **Originality.** By the said parameters, the two-layer perceptron is trained optimally. Its performance over turned objects with 4800 features is nearly the best. The mentioned quasioptimality is unavoidable because of very high variance at the lower classification error percentage. **Practical value.** The averaged classification error percentage decreases down to 0,13 %. At the maximal intensity of turn distortion, the classification error percentage decreases down to 0,85 %. References 10, figures 5.

Key words: turn distortion, classification, perceptron, training, performance, classification error percentage, optimization.

ОПТИМАЛЬНЕ ЧИСЛО НЕЙРОНІВ ПРИХОВАНОГО ШАРУ ДВОШАРОВОГО ПЕРСЕПТРОНУ ТА СПІВВІДНОШЕННЯ СЕРЕДНЬОКВАДРАТИЧНИХ ВІДХИЛЕНЬ ДЛЯ ЙОГО НАВЧАННЯ НА ПОВЕРНУТИХ 60×80-ЗОБРАЖЕННЯХ З ПІКСЕЛЬНИМИ СПОТВОРЕННЯМИ В КЛАСИФІКАЦІЇ ОБ'ЄКТІВ З ПОВОРОТАМИ

В. В. Романюк

Хмельницький національний університет

вул. Інститутська, 11, м. Хмельницький, 29000, Україна. E-mail: romanukevadimv@mail.ru

Оптимізується двошаровий персептрон за його двома параметрами для задачі класифікації об'єктів з поворотами. Першим параметром є число нейронів у прихованому шарі. Другим параметром є співвідношення середньоквадратичних відхилень піксельних спотворень і поворотів при формуванні навчальної множини. Знайдено оптимальне число нейронів у прихованому шарі двошарового персептрона поряд зі співвідношенням середньоквадратичних відхилень піксельних спотворень і поворотів для навчання на повернутих 60×80 зображеннях з піксельними спотвореннями. Пара таких оптимальних параметрів, втім, не є єдиною, і наводяться інші варіанти оптимальних пар. Двошаровий персептрон за цими параметрами навчається в оптимальному режимі. Його продуктивність на повернутих об'єктах з 4800 ознаками є майже найкращою. Середній відсоток помилок класифікації зменшується до 0,13 %. За максимальної інтенсивності спотворень поворотами відсоток помилок класифікації зменшується до 0,85 %.

Ключові слова: спотворення поворотами, класифікація, персептрон, навчання, продуктивність, відсоток помилок класифікації, оптимізація.

PROBLEM STATEMENT. *Twists, angulations, swerves in objects and their classification*. Classification is a foundation for automatization and control. Before constructing a classification system, objects for classification are considered with their features and distortions in them, what is needed to identify the classifier [1, 2]. These distortions are twists, angulations, swerves, which are as frequent as shifts, scalings, feature distortions [3, 4].

Once the number of classes N_0 is determined within the general totality (GT), containing non-distorted pure representatives (NDPR) of all these classes (one NDPR for each class), the classifier identification starts. In classifying objects with their properties of turn (rotation or twist, angulation, swerve), the number N_0 depends on how hard the object may be turned. The greater the maximal turn angle (MTA) $\alpha_{\max}$ is, the greater N_0 should be. For instance, if $\alpha_{\max} = \pi/2$ then it is advisable to differ NDPR of the plane object, turned by a quarter of full turn, and the initial NDPR of this object (without any distortion, without turn). Thus N_0 is doubled. If $\alpha_{\max} = 2\pi$ in plane objects classification, that is the object can be turned arbitrarily around its center, then it is worthwhile to put into 16 different classes the same object, turned at angle $\alpha_n = n\pi/8$ by $n = \overline{0, 15}$ (figure 1).

Another method of classifying arbitrarily turned objects consists in using 16 classifiers for objects, whose NDPR originate from N_0 initial NDPR with the angle α_0 (non-turned), subsequently turned at angles in the set $\{\alpha_n = n\pi/8\}_{n=0}^{15}$. Generally, there should be used

$2M$ classifiers if the $(n+1)$-th class object NDPR is the initial (non-turned) NDPR of this object, turned at angle

$$\alpha_n = \frac{n\pi}{M} \text{ by } n = \overline{0, 2M-1} \text{ for } M \in \mathbb{N}. \quad (1)$$

Figure 1 – The plane monochrome imaged object, turned at angle $\alpha_n = n\pi/8$ and put into the class $n+1$ by $n = \overline{0, 15}$

This clearly lingers the classification process. The division by $2M$ classifiers with N_0 classes for each classifier is accomplished when the objects data are preprocessed [2, 5, 6], what slightly accelerates the classification process in comparison to using the single classifier with the increased number of classes $2MN_0$ instead of the initial N_0.

Improving the classification process with the classifier parameters optimization. Suppose that the classifier has $K \in \mathbb{N}$ adjustable parameters $\left\{a_p\right\}_{p=1}^{K}$ by $a_p \in A_p$ $\forall\ p = \overline{1, K}$ and A_p is the set of admissible values of the parameter a_p. There may be sets of linguistic variables in $\left\{A_p\right\}_{p=1}^{K}$.

The function of performance $\vartheta\left(\left\{a_p\right\}_{p=1}^{K}\right)$ is an aggregate over output parameters of the classifier — classification error percentage (CEP), traintime duration (TTD), hang probability, level of resources (memory) consumption, etc. Usually, the aggregate performance is a classifier property to be minimized. If

$$\mathbf{A} = \left[a_p\right]_{1\times K} \in \bigtimes_{p=1}^{K} A_p$$

then applying the vector

$$\mathbf{A}^* = \left[a_p^*\right]_{1\times K} \in \arg \min_{\mathbf{A} \in \bigtimes_{p=1}^{K} A_p} \vartheta\left(\left\{a_p\right\}_{p=1}^{K}\right) \quad (2)$$

of optimal parameters for the classifier improves its functionality and the classification process on the whole.

Some of those parameters relate to the classifier structure, others are included into training sets for supervised learning process (SLP). Forming the training sets for SLP rationally is very important to identify the classifier for the given GT in the best way [1]. However, for solving the problem (2), there must be evaluated the sets $\left\{A_p\right\}_{p=1}^{K}$ (commonly, with their boundaries) and the function of performance $\vartheta\left(\left\{a_p\right\}_{p=1}^{K}\right)$. This is pretty robust work on multivariate statistics, which gets complicated as the quantity K increases [2, 6].

Thus discourse on defining and solving the problem (2) is reasonable when classifiers have short TTD in SLP and consume not much of resources. Perceptrons are good in those two aspects. Theoretically, two-layer perceptron (2LP) can fit any function, which maps the being investigated objects into their N_0 classes [7]. For that 2LP structure ought to be rationalized by adjusting its hidden layer neurons number (HLNN) N_{HLN}. In turned objects classification problem, 2LP has better to be trained on turned objects, mixed with additional feature distortion (AFD). For plane objects, imaged monochrome, this AFD is pixel distortion (PD), when the pixel $\{u, v\}$ code (value) $\upsilon_{uv} \in \{0, 1\}$ of the image formatted $U \times V$ is converted to a real value $\tilde{\upsilon}_{uv} \notin \{0, 1\}$ at $u \in \left\{\overline{1, U}\right\}$ and $v \in \left\{\overline{1, V}\right\}$. In degenerate cases, PD is pixel inversion due to $\tilde{\upsilon}_{uv} = 1 - \upsilon_{uv}$.

The intensity of PD can be measured (or preset) with standard deviation (SD) σ_{PD} [3]. If there is SD $\hat{\sigma}$ that defines the intensity of turn distortion (MTA $\alpha_{\max}$), then pixel-to-turn standard deviations ratio (PTSDR)

$$r = \sigma_{\text{PD}} / \hat{\sigma} \quad (3)$$

stands as one of the training process parameters from the set $\{a_p\}_{p=1}^{K}$. The questions of this paper are the following. What is the most rational (optimal) HLNN for 2LP, classifying turned 60×80 monochrome images (T6080MI)? And at what PTSDR shall 2LP be trained to obtain its performance with lower CEP? Clearly, these two parameters of 2LP are going to be optimized simultaneously alike in (2).

The paper goal and tasks. Let 2LP performance function $\vartheta(r, N_{\text{HLN}})$ be an averaged CEP value $p_{\text{EC}}(r, N_{\text{HLN}})$. Then the problem (2) is

$$\begin{bmatrix} r^* & N^*_{\text{HLN}} \end{bmatrix} \in \arg\min_{A_1\times A_2} p_{\text{EC}}(r, N_{\text{HLN}}) \qquad (4)$$

implicitly. For solving the problem (4) there are GT and NDPR of the classes to be defined, whereupon a method to train 2LP will be accepted. Based on a model of pixel-distorted turned 60×80 monochrome images (PDT6080MI), boundaries for HLNN and boundaries for PTSDR are to be evaluated that the global minimum of HLNN and the global minimum of PTSDR would be certainly enclosed within the corresponding ranges of HLNN and PTSDR. The next stage is running through the rectangle of HLNN and PTSDR values, what will let minimize the averaged CEP due to (4). Inferences and suggestions for further 2LP performance optimization will conclude this work.

EXPERIMENTAL PART AND RESULTS OBTAINED. *GT and NDPR of the classes.* If a plane image is monochrome then it is modeled with 60×80 matrix of ones and zeros (MOZ). Hence GT of all monochrome 60×80 images is finite and has 2^{4800} elements, 2^{4800} those MOZ. To embody the plane monochrome imaged object, NDPR of N_0 classes within this GT should have some generalized properties like horizontal and vertical lines, squares, circles, slants, curves or serpentine lines. The enlarged English alphabet capital letter (EEACL) could be assumed for that (there are downsized monochrome 60×80 images of all 26 EEACL in figure 1). There won't be considered angulations greater than a quarter of full turn, therefore $N_0 = 26$.

MATLAB function for training the perceptron. A lot of methods to train 2LP are developed within MATLAB Neural Network Toolbox [3, 4]. Basically, they are grounded on backpropagation algorithm. Statistically the best convergence in SLP is provided with the training MATLAB function "traingda" [3, 8, 9], which updates weight and bias values according to gradient descent with adaptive learning rate [10]. Henceforward, "traingda" will be used in SLP to identify the classifier on the ground of 2LP, which in MATLAB is initialized as the function "feedforwardnet" (or "newff", in obsolete way). Let this 2LP be denoted by $\mathcal{P}_0(4800, N_{\text{HLN}}, 26)$ notifying that 2LP input layer has 4800 neurons, its SHL has N_{HLN} neurons, and 2LP output layer has $N_0 = 26$ neurons.

Model of PDT6080MI. The training set with PDT6080MI is formed in $F\in\mathbb{N}$ phases. Number F is an indicator of smoothness in SLP of $\mathcal{P}_0(4800, N_{\text{HLN}}, 26)$. The greater its value is, the smoother $\mathcal{P}_0(4800, N_{\text{HLN}}, 26)$ is trained, acquiring better performance properties. However, increasing exceedingly the number F delays SLP. That's why $F = 8$ what is far enough for $\mathcal{P}_0(4800, N_{\text{HLN}}, 26)$.

SD, defining the intensity of turn distortion at the k-th phase, is

$$\hat{\sigma}_k = \hat{\sigma}_{\max}\cdot F^{-1}\cdot k \quad \forall\, k=\overline{1, F} \qquad (5)$$

by $\hat{\sigma}_{\max} > 0$. Every NDPR from GT is turned at angle

$$\beta(k) = \frac{180}{\pi}\cdot\hat{\sigma}_k\xi(k) \qquad (6)$$

in degrees around the center point of the image, where $\xi(k)$ is a value of normal variate (NV) with zero expectation and unit variance (ZEUV), raffled at the k-th phase. The q-th class NDPR as 60×80 MOZ $\mathbf{H}_q = \left(h_{uv}^{\langle q\rangle}\right)_{60\times80}$ is processed into T6080MI

$$\hat{\mathbf{H}}_q(k) = 1 - \boldsymbol{\rho}\left(1-\mathbf{H}_q, \beta(k), M, S\right) \ \text{ at } \ q=\overline{1, 26} \quad (7)$$

by the map $\boldsymbol{\rho}$, turning the input q-th class NDPR negative $1-\mathbf{H}_q$ at angle (6), where the interpolation method with the handle M is applied, and the size 60×80 of the returned negative T6080MI of the q-th class is specified with the handle S [3]. The map $\boldsymbol{\rho}$ in (7) along with handles $\{M, S\}$ is implemented by MATLAB function "imrotate" [4]. NDPR is turned in counterclockwise direction if $\beta(k) > 0$, and for $\beta(k) < 0$ NDPR is turned clockwise; of course, NDPR remains itself for $\beta(k) = 0$.

At the k-th phase PDT6080MI is 4800×26 matrix

$$\tilde{\hat{\mathbf{H}}}(k) = \hat{\mathbf{H}}(k) + \sigma_{\text{PD}}^{\langle k\rangle}\cdot\boldsymbol{\Xi} \qquad (8)$$

by SD

$$\sigma_{\text{PD}}^{\langle k\rangle} = \sigma_{\text{PD}}^{\langle\max\rangle}\cdot F^{-1}\cdot k \quad \forall\, k=\overline{1, F} \qquad (9)$$

and matrix $\hat{\mathbf{H}}(k) = \left[\bar{\hat{h}}_{jq}(k)\right]_{4800\times26}$ whose q-th column is 4800-length-single-column-reshaped matrix (4800-LSCRM) $\hat{\mathbf{H}}_q(k)$ in (7), where $\boldsymbol{\Xi}$ is 4800×26 matrix of values of NV with ZEUV. Maximal intensity of PD $\sigma_{\text{PD}}^{\langle\max\rangle} > 0$ is connected to the similar one $\hat{\sigma}_{\max} > 0$ in (5) across PTSDR (3). This PTSDR is constant $\forall\, k=\overline{1, F}$ and so there is more strict statement

$$r = \sigma_{\text{PD}}^{\langle\max\rangle} \big/ \hat{\sigma}_{\max} \qquad (10)$$

for PTSDR. PTSDR (10) varies as $\sigma_{\text{PD}}^{\langle\max\rangle}$ changes on some constant $\hat{\sigma}_{\max}$, being definition-maker for T6080MI. Henceforward, for presetting PTSDR there is SD $\sigma_{\text{PD}}^{\langle\max\rangle}$ is just required.

For training on PDT6080MI the input of $\mathcal{P}_0(4800, N_{\text{HLN}}, 26)$ is fed with the training set

$$\left\{\{\mathbf{H}\}_{d=1}^{R}, \left\{\tilde{\hat{\mathbf{H}}}(k)\right\}_{k=1}^{F}\right\} \qquad (11)$$

by $R\in\mathbb{N}$ replicas of all 26 classes NDPR, where the q-th column of 4800×26 matrix $\mathbf{H}$ is

4800-LSCRM $\mathbf{H}_q$ at $q=\overline{1, 26}$. Number of replicas of $\mathbf{H}$ slightly influences on SLP smoothness, and $R=2$ is far enough for $\mathscr{P}_0(4800, N_{\text{HLN}}, 26)$. The set (11), being $4800 \times 26(R+F)$ matrix, is passed through $\mathscr{P}_0(4800, N_{\text{HLN}}, 26)$ with $R+F$ identifiers $\{\mathbf{I}\}_{i=1}^{R+F}$ on identity 26×26 matrix $\mathbf{I}$ for Q_{pass} times. Thus the PDT6080MI-trained $\mathscr{P}_0(4800, N_{\text{HLN}}, 26)$ can be denoted by

$$\mathscr{P}\left(4800, N_{\text{HLN}}, 26; \widehat{\sigma}_{\max}, r; R, F, Q_{\text{pass}}\right). \quad (12)$$

To obtain

$$\mathscr{P}\left(4800, N_{\text{HLN}}, 26; \widehat{\sigma}_{\max}, r; 2, 8, Q_{\text{pass}}\right)$$

with high performance quality in classifying T6080MI or PDT6080MI, it is required to preset $Q_{\text{pass}} \geqslant 75$. Therefore, after fixing some $\widehat{\sigma}_{\max}$,

$$\mathscr{P}\left(4800, N_{\text{HLN}}, 26; \widehat{\sigma}_{\max}, r; 2, 8, 75\right) \quad (13)$$

will be tested as the classifier of T6080MI and PDT6080MI through the line (or striped, because HLNN are integers) rectangle of HLNN and PTSDR values.

Boundaries for HLNN. The range of HLNN values is $\left[N_{\text{HLN}}^{\langle\min\rangle}; N_{\text{HLN}}^{\langle\max\rangle}\right] \cap \mathbb{N}$ by its boundaries $N_{\text{HLN}}^{\langle\min\rangle}$ and $N_{\text{HLN}}^{\langle\max\rangle}$. Left boundary $N_{\text{HLN}}^{\langle\min\rangle}$ is that by $N_{\text{HLN}} < N_{\text{HLN}}^{\langle\min\rangle}$ 2LP $\mathscr{P}_0(4800, N_{\text{HLN}}, 26)$ is trained too slow or cannot be trained at all. Right boundary $N_{\text{HLN}}^{\langle\max\rangle}$ is that by $N_{\text{HLN}} > N_{\text{HLN}}^{\langle\max\rangle}$ 2LP $\mathscr{P}_0(4800, N_{\text{HLN}}, 26)$ can be over-trained or SLP hangs and buzzes even while it is trained on NDPR with their 4800×26 matrix $\mathbf{H}$. Empirically, there are most likely $N_{\text{HLN}}^{\langle\min\rangle} = 150$ and $N_{\text{HLN}}^{\langle\max\rangle} = 400$ independently of PTSDR (10) for sensible SD $\widehat{\sigma}_{\max}$. Now, 2LP (13) shall be run through batch testing by $N_{\text{HLN}} \in [150; 400] \cap \mathbb{N}$ under some $\widehat{\sigma}_{\max}$ and PTSDR.

Boundaries for PTSDR. It is clear that $\widehat{\sigma}_{\max}$ can be assigned according to the restriction on angulations for T6080MI within a quarter of full turn. Again empirically, with being generated angles (6) value $\widehat{\sigma}_{\max} = 0.2$ is suitable for that restriction [3]. The lowest value of $\sigma_{\text{PD}}^{\langle\max\rangle}$ cannot be zero, inasmuch as at $r=0$ there is no PDT6080MI (8), $\mathscr{P}_0(4800, N_{\text{HLN}}, 26)$ is trained solely on F portions $\left\{\tilde{\tilde{\mathbf{H}}}(k)\right\}_{k=1}^{F}$ of T6080MI, and SLP is delayed unwantedly. Besides, 2LP (13) should classify PDT6080MI, having admittedly its own PTSDR for testings (it'll be argued against in the next section). This is workable if $r>0$. Thereupon the lowest value of $\sigma_{\text{PD}}^{\langle\max\rangle}$ is 0.002 as by $\sigma_{\text{PD}}^{\langle\max\rangle} < 0.002$ PD in PDT6080MI (8) is hardly tangible, and the uppermost value of $\sigma_{\text{PD}}^{\langle\max\rangle}$ is 0.2 as by $\sigma_{\text{PD}}^{\langle\max\rangle} > 0.2$ PDT6080MI (8) become over-distorted (visually they are perceived as pure grayed noise). Consequently, range $\left[r_{\min}; r_{\max}\right]$ of PTSDR (10) has boundaries $r_{\min} = 0.01$ and $r_{\max} = 1$.

Running through the rectangle of HLNN and PTSDR. The line rectangle of HLNN and PTSDR values

$$\left[r_{\min}; r_{\max}\right] \times \left\{\left[N_{\text{HLN}}^{\langle\min\rangle}; N_{\text{HLN}}^{\langle\max\rangle}\right] \cap \mathbb{N}\right\} = \\ = [0.01; 1] \times \left\{[150; 400] \cap \mathbb{N}\right\} \quad (14)$$

is to be covered with evaluations of the averaged CEP values $p_{\text{EC}}(r, N_{\text{HLN}})$ for solving the problem (4)

$$\left[r^* \quad N_{\text{HLN}}^*\right] \in \\ \in \arg\left(\min_{\left[r \quad N_{\text{HLN}}\right] \in [0.01; 1] \times \{[150; 400] \cap \mathbb{N}\}} \left\{p_{\text{EC}}(r, N_{\text{HLN}})\right\}\right). \quad (15)$$

For solving the problem (15), 2LP

$$\mathscr{P}\left(4800, N_{\text{HLN}}, 26; 0.2, r; 2, 8, 75\right) \quad (16)$$

shall be run through batch testing by

$$\left[r \quad N_{\text{HLN}}\right] \in [0.01; 1] \times \left\{[150; 400] \cap \mathbb{N}\right\}. \quad (17)$$

The batch testing implies that 2LP (16) is tested with T6080MI at

$$\widehat{\sigma} \in \left[0; \widehat{\sigma}_{\max}\right] = [0; 0.2]$$

and with PDT6080MI under

$$\widehat{\sigma} \in \left(0; \widehat{\sigma}_{\max}\right] = (0; 0.2]$$

at the testing PTSDR $r_{\text{test}}(\widehat{\sigma}) \in (0; 1]$ by (3). The uppermost value of $r_{\text{test}}(\widehat{\sigma})$ has been suggested for AFD wouldn't be more than just the turn intensity. Stating it generally, at an SD $\widehat{\sigma} \in \boldsymbol{S}$ by a distortion type $\boldsymbol{\tau} \in \boldsymbol{\Theta}$ and a testing PTSDR $r_{\text{test}}(\widehat{\sigma}) \in \boldsymbol{R}(\widehat{\sigma})$ 2LP (16) produces CEP

$$p_{\text{EC}}\left(r, N_{\text{HLN}}, \widehat{\sigma}, \boldsymbol{\tau}, r_{\text{test}}(\widehat{\sigma})\right) \quad (18)$$

at (17). In CEP (18) $\boldsymbol{S}$ is a range of SD and $\boldsymbol{R}(\widehat{\sigma})$ is a range of the testing PTSDR by a fixed point from $\boldsymbol{S}$ (here, the range of the testing PTSDR won't be varied). The set $\boldsymbol{\Theta}$ is either of linguistic variable values, labeling distortion types (like values "T6080MI", "PDT6080MI", etc.), or of numeric values. Reasonably to put it $\boldsymbol{\Theta} \subset \mathbb{N}$ and then $\boldsymbol{\tau} = 1$ means "T6080MI", $\boldsymbol{\tau} = 2$ means "PDT6080MI", $\boldsymbol{\tau} = 3$ could have meant some else distortion, engaging SD $\widehat{\sigma} \in \boldsymbol{S}$. For $\boldsymbol{\tau} = 1$ the testing PTSDR $r_{\text{test}}(\widehat{\sigma}) = 0 \quad \forall\, \widehat{\sigma} \in \boldsymbol{S}$, for $\boldsymbol{\tau} = 2$ the testing PTSDR $r_{\text{test}}(0) = 0$. Then that averaged CEP value under minimization operator in (15) is a convex combination

$$p_{\text{EC}}(r, N_{\text{HLN}}) = \lambda(1)\left(\frac{1}{\|\boldsymbol{S}\|}\int_{\widehat{\sigma} \in \boldsymbol{S}} p_{\text{EC}}\left(r, N_{\text{HLN}}, \widehat{\sigma}, 1, 0\right) d\widehat{\sigma}\right) +$$

$$+\lambda(2)\left(\frac{1}{\|\boldsymbol{S}\|}\int_{\hat{\sigma}\in\boldsymbol{S}}\left[\frac{1}{\|\boldsymbol{R}(\hat{\sigma})\|}\int_{r_{\text{test}}(\hat{\sigma})\in\boldsymbol{R}(\hat{\sigma})} p_{\text{EC}}\left(r,\, N_{\text{HLN}},\, \hat{\sigma},\, 2,\, r_{\text{test}}(\hat{\sigma})\right)dr_{\text{test}}(\hat{\sigma})\right]d\hat{\sigma}\right) \qquad (19)$$

by $\lambda(1)>0$, $\lambda(2)>0$, $\lambda(1)+\lambda(2)=1$. In generality,

$$\lambda(\boldsymbol{\tau})>0 \quad \forall\, \boldsymbol{\tau}\in\boldsymbol{\Theta} \quad \text{and} \quad \sum_{\boldsymbol{\tau}\in\boldsymbol{\Theta}}\lambda(\boldsymbol{\tau})=1\,, \qquad (20)$$

where coefficient $\lambda(\boldsymbol{\tau})$ in (20) for the convex combination in (19) designates involvement of the distortion type $\boldsymbol{\tau}$ into the averaged CEP value. If the set $\boldsymbol{S}$ includes the point $\hat{\sigma}=0$, in which 2LP works on NDPR, then the second integral in (19) is brought to zero according to the norm $\|\boldsymbol{R}(\hat{\sigma})\|>0 \;\; \forall\, \hat{\sigma}\in\boldsymbol{S}$. On local simplification, combination (19) is re-stated as

$$p_{\text{EC}}(r,\, N_{\text{HLN}})=\lambda(1)\left(\frac{1}{0.2-0}\int_{\hat{\sigma}\in[0;\,0.2]} p_{\text{EC}}(r,\, N_{\text{HLN}},\, \hat{\sigma},\, 1,\, 0)\,d\hat{\sigma}\right)+$$

$$+\lambda(2)\left(\frac{1}{0.2-0}\int_{\hat{\sigma}\in(0;\,0.2]}\left[\frac{1}{1-0}\int_{r_{\text{test}}(\hat{\sigma})\in(0;\,1]} p_{\text{EC}}\left(r,\, N_{\text{HLN}},\, \hat{\sigma},\, 2,\, r_{\text{test}}(\hat{\sigma})\right)dr_{\text{test}}(\hat{\sigma})\right]d\hat{\sigma}\right)=$$

$$=5\lambda(1)\int_{\hat{\sigma}\in[0;\,0.2]} p_{\text{EC}}(r,\, N_{\text{HLN}},\, \hat{\sigma},\, 1,\, 0)\,d\hat{\sigma}+5\lambda(2)\int_{\hat{\sigma}\in(0;\,0.2]}\left[\int_{r_{\text{test}}(\hat{\sigma})\in(0;\,1]} p_{\text{EC}}\left(r,\, N_{\text{HLN}},\, \hat{\sigma},\, 2,\, r_{\text{test}}(\hat{\sigma})\right)dr_{\text{test}}(\hat{\sigma})\right]d\hat{\sigma}\,. \qquad (21)$$

Ranges $[0;\, 0.2]$ and $(0;\, 0.2]$ are going to be run through with the step 0.02 what lets evaluate the averaged CEP (21) numerically:

$$p_{\text{EC}}(r,\, N_{\text{HLN}})=\frac{\lambda(1)}{11}\sum_{s=0}^{10} p_{\text{EC}}(r,\, N_{\text{HLN}},\, 0.02s,\, 1,\, 0)+$$

$$+\frac{\lambda(2)}{100}\sum_{s=1}^{10}\sum_{w=1}^{10} p_{\text{EC}}(r,\, N_{\text{HLN}},\, 0.02s,\, 2,\, 0.1w) \qquad (22)$$

for each two-component point (17), where norms $\|\boldsymbol{S}\|$ and $\|\boldsymbol{R}(\hat{\sigma})\|$ have been substituted with cardinalities of the sampled subsets

$$\{0.02s\}_{s=0}^{10}\subset[0;\, 0.2],\;\; \{0.02s\}_{s=1}^{10}\subset(0;\, 0.2],$$
$$\{0.1w\}_{w=1}^{10}\subset(0;\, 1].$$

Involvement of T6080MI and PDT6080MI could be considered closely equal. And it might have been put $\lambda(1)=\lambda(2)=0.5$ in (22). However, testing with PDT6080MI takes longer periods, and it does not correct evaluations of the averaged CEP much (see, for instance, [3, 4]).

The rectangle (14) is sampled differently from sides of PTSDR and HLNN. For the first time, HLNN is taken with some gaps:

$$N_{\text{HLN}}\in\left\{150,\, \{200+10q\}_{q=0}^{15},\, 400\right\}.$$

The range (side) $[0.01;\, 1]$ of PTSDR is sampled into the discrete subrange

$$\left\{\{0.01+0.01l\}_{l=0}^{8},\, \{0.1l\}_{l=1}^{10}\right\}\subset[0.01;\, 1].$$

The corresponding 342 sampled estimation values of CEP

$$p_{\text{EC}}(r,\, N_{\text{HLN}})=\frac{1}{11}\sum_{s=0}^{10} p_{\text{EC}}(r,\, N_{\text{HLN}},\, 0.02s,\, 1,\, 0) \qquad (23)$$

on 400 batch testings of 2LP (16) allow to see (figure 2) a rough mesh (23) of the surface (21) above the point rectangle (where each point is the average of seven 2LP)

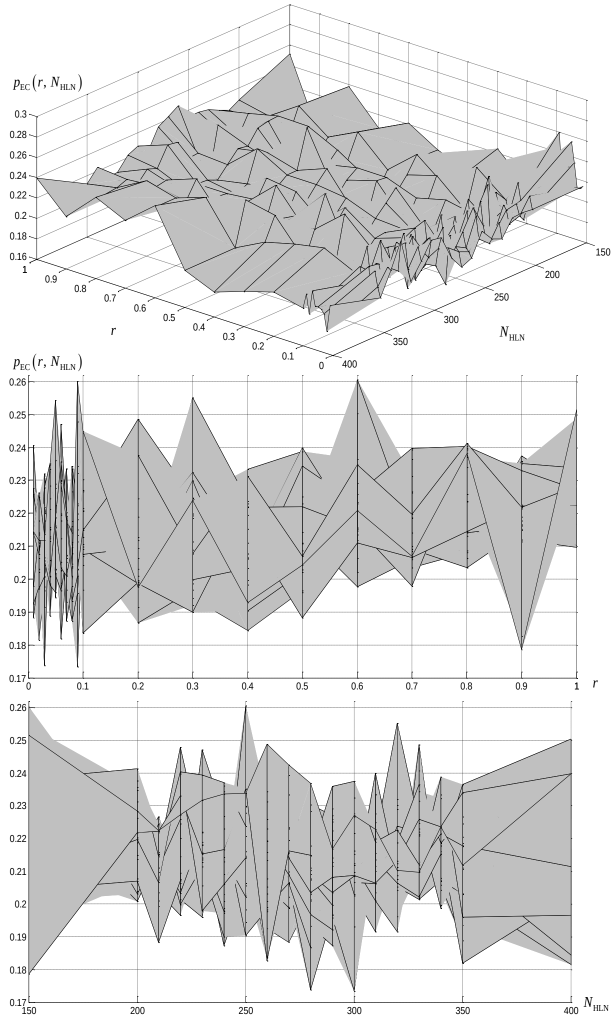

Figure 2 – A rough mesh (23) of the surface (21) by $\lambda(1)=1$ above the point rectangle (24)

$$\left\{\left\{0.01+0.01l\right\}_{l=0}^{8},\ \left\{0.1l\right\}_{l=1}^{10}\right\}\times$$
$$\times\left\{150,\ \left\{200+10q\right\}_{q=0}^{15},\ 400\right\}. \qquad (24)$$

The rough-watched minima of the surface (21) allow narrowing the point rectangle (24) to 238-pointed one

$$\left\{\left\{0.01+0.01l\right\}_{l=0}^{8},\ \left\{0.1l\right\}_{l=1}^{5}\right\}\times\left\{\left\{200+10q\right\}_{q=0}^{15},\ 400\right\}. \quad (25)$$

Now the refined mesh (23) of the surface (21) above the point rectangle (24) is seen (figure 3) more scrupulous, where each point is the average of 10 2LP. Further, HLNN 360 is included, and PTSDR greater than 0.1 are cut off. Figure 4 represents the locally re-refined mesh (23) of the surface (21) above the 170-pointed rectangle

$$\left\{\left\{0.01l\right\}_{l=1}^{10}\right\}\times\left\{\left\{200+10q\right\}_{q=0}^{16}\right\}. \qquad (26)$$

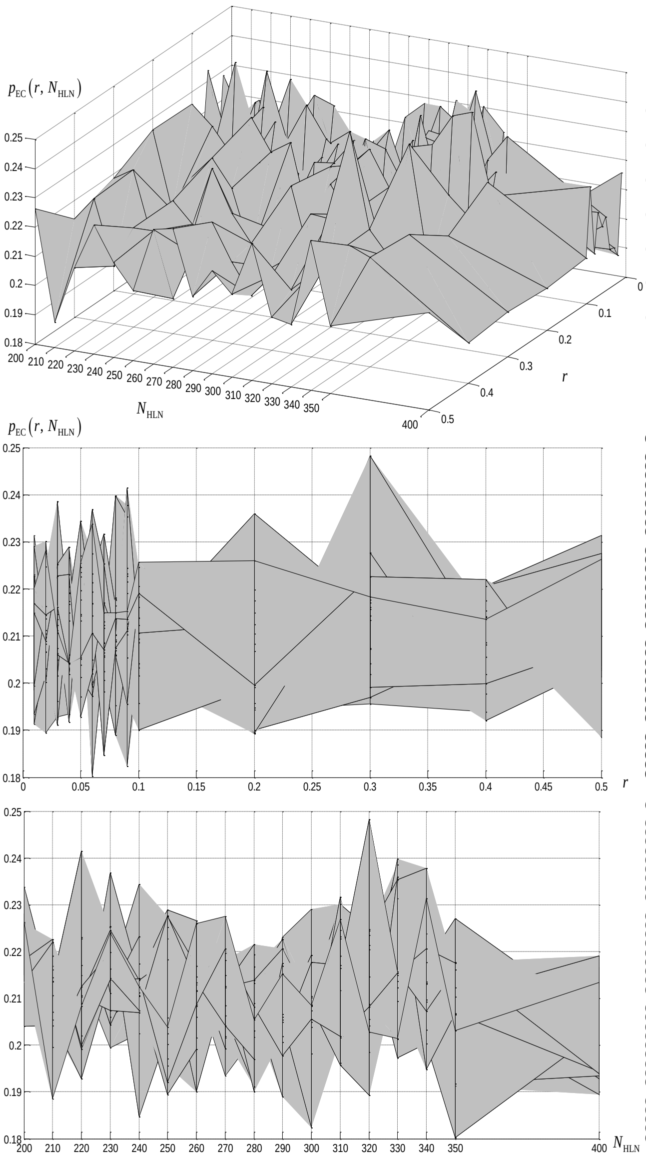

Figure 3 – The locally refined mesh (23) of the surface (21) by $\lambda(1)=1$ above the point rectangle (25)

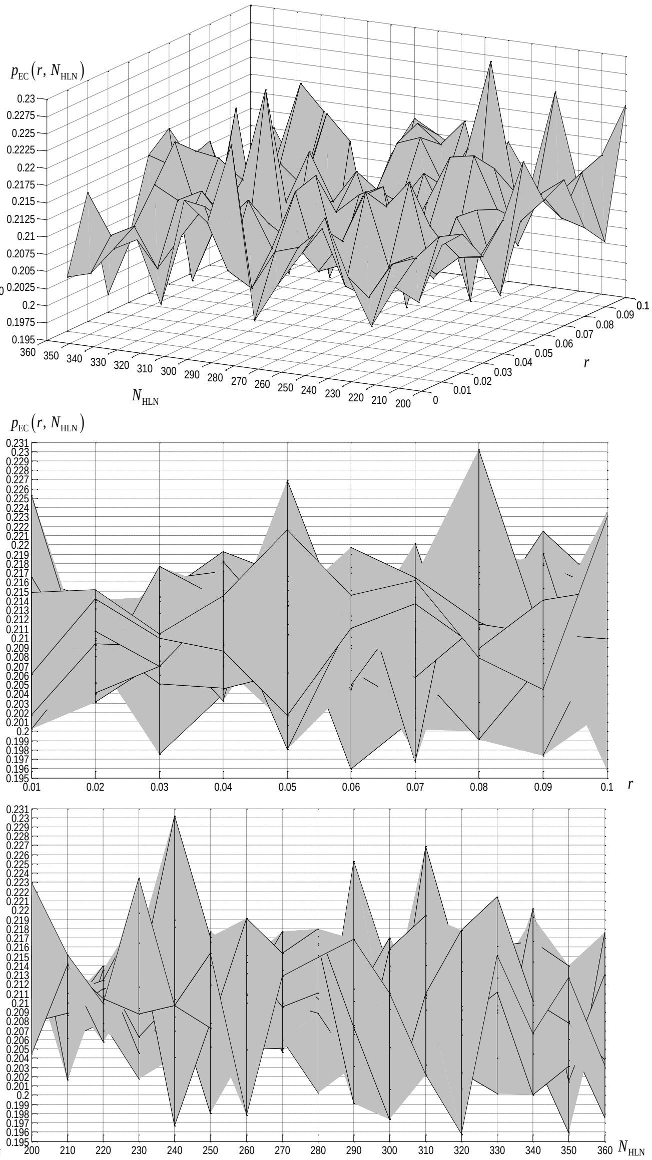

Figure 4 – The locally re-refined mesh (23) of the surface (21) by $\lambda(1)=1$ above the point rectangle (26), where each point is the average of 35 2LP

Need we further narrowing the 170-pointed rectangle (26)? For instance, it might have been taken for

$$\left\{\left\{0.05+0.01l\right\}_{l=0}^{5}\right\}\times\left\{\left\{220+10q\right\}_{q=0}^{10}\right\}.$$

Proceeding from those 35 2LP used for averaging over badly volatile local evaluations, the answer is negative: CEP is too small and it is at about the variance. This is why taking HLNN with the unit step is senseless here.

The averaged CEP minimization due to the problem (15) and its verification. After having analyzed the surface (21) meshes for T6080MI classification problem, the problem (15) numerical solution looks like to be in a series of untied points, whose CEP is less than 0.2 (figure 5). The smallest CEP is at the point

$$\left[r \quad N_{\mathrm{HLN}}\right]=\left[0.1 \quad 320\right], \qquad (27)$$

but the point

$$\left[r \quad N_{\mathrm{HLN}}\right]=\left[0.06 \quad 350\right] \qquad (28)$$

is much the same. However, if to consider all

$170 \cdot 35 = 5950$ 2LP for figure 4, there are 16 2LP (contoured elliptically and highlighted in figure 5) whose CEP occurred to be less than 0.13, although points (27) and (28) don't belong to the list of those 16 ones. Hence, the best 2LP is searched within this list.

N_{HLN} \ r	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	**0.1**
200	0.2149	0.2152	0.2105	0.2145	0.2216	0.2146	0.2162	0.2078	0.2044	0.2229
210	0.2061	0.2142	0.2100	0.2086	0.2016	0.2110	0.2137	0.2088	0.2140	0.2152
220	0.2078	0.2108	0.2070	0.2140	0.2136	0.2124	0.2057	0.2115	0.2104*	0.2099
230	0.2017	0.2094	0.2091	0.2045	0.2063	0.2197	0.2165	0.2117	0.2087	0.2235
240	0.2084	0.2041	0.2070*	0.2182	0.2104	0.2189	*0.1967*	0.2302	0.2097	0.2097
250	0.2166	0.2052	0.2177	0.2140	*0.1980*	0.2120	0.2078	0.2171	0.2072	0.2153
260	0.2108	0.2132	0.2100	0.2151	0.2151	0.2049	0.2143	0.2110	0.2191	*0.1978*
270	0.2059	0.2100	0.2051	0.2046	0.2136	0.2051	0.2095	0.2177	0.2154	0.2128
280	0.2002	0.2089	0.2162	0.2172	0.2105	0.2151	0.2110	0.2164	0.2180	0.2149
290	0.2253	0.2031	0.2076	0.2071	0.2115	0.2102	0.2066	*0.1991*	0.2073	0.2168
300	0.2135	0.2052	0.2140	0.2093	0.2006	0.2170	*0.1975*	0.2158	*0.1974*	0.2110
310	0.2164	0.2141	0.2105	0.2032	0.2269	0.2092	0.2108	0.2194	0.2109	0.2022
320	0.2007	0.2092	0.2144	0.2096*	0.2134	0.2065	0.2024	0.2082	0.2179	*0.1958*
330	0.2097	0.2040	0.2127	0.2092	0.2161	0.2089*	0.2002	0.2111	0.2215	0.2151
340	0.2079	0.2098	0.2145	0.2193	0.2166	0.2044	0.2202	0.2000	0.2102	0.2067
350	0.2038	0.2080	0.2061	0.2080	0.2140	*0.1960*	0.2014	0.2031	0.2078	0.2127
360	0.2028	0.2138	*0.1975*	0.2039	0.2151	0.2176	0.2118	0.2131	0.2038	0.2032

Figure 5 – CEP values of the mesh (23) above the point rectangle (26), where nine less-than-0.2 values are highlighted

It is revealed that 2LP is trained best at

$$\left[r^* \quad N^*_{\text{HLN}} \right] = [0.04 \quad 320]. \tag{29}$$

In this point the averaged-rounded-upward CEP

$$p_{\text{EC}}\left(r^*, N^*_{\text{HLN}}\right) = p_{\text{EC}}(0.04, 320) = 0.13$$

appears to be minimal within the line rectangle (14). Is it really minimal and does the point (29) indeed minimize the averaged CEP over T6080MI, tolerating some PDT6080MI? Yes, this is successfully verified with 1000 batch testings. Noteworthy, at the maximal intensity of turn distortion, i. e. at $\hat{\sigma} = \hat{\sigma}_{\max} = 0.2$, CEP over T6080MI decreases down to 0.85 %. This confirms that the point (29) is a solution of the problem (15). Nonetheless, this solution is not unique, and other optimal (or, rather, quasioptimal) pairs of PTSDR and HLNN are right-above-starred in figure 5.

CONCLUSIONS. The mentioned quasioptimality is understood clearer when watching meshes in figure 4 and signs in figure 5. This "quasi" is unavoidable because of very high variance at the lower CEP. The most rational HLNN for 2LP to classify turned objects with 4800 features lies between 220 and 330. An exact number cannot be specified, though an HLNN from 320 to 330 is strongly recommended. A quasioptimal PTSDR should be selected from the segment $[0.03; 0.1]$, especially if number of features is comparable to a few thousands, but it's not 4800. For T6080MI or PDT6080MI, the most appropriate PTSDR is a value from the segment $[0.04; 0.06]$. Generally, for classifying turned objects, HLNN shall be adjusted to the set $\left\{ \overline{220, 350} \right\}$ of 131 elements (see this range in figure 5 vertically), and the best PTSDR is $r^* \in [0.03; 0.1]$, although roughly.

Inferences and suggestions for further 2LP performance optimization. The described procedure of optimizing the performance of 2LP (12) in its two parameters for a turned objects classification problem can be applied to other 2LP

$$\mathscr{P}\left(N_{\text{format}}, N_{\text{HLN}}, N_0; \mathfrak{T}\right) \tag{30}$$

with the sequence $\mathfrak{T}$ of SLP parameters, when 2LP (30) has to classify objects with their N_{format} features. The structure of this paper is a template for such optimization. The prime stage is to define GT and NDPR of the classes in it, whereupon 2LP and turn objects distortion are modeled. Subsequent stages are: to evaluate boundaries for 2LP parameters, which are to be optimized; to run 2LP through the set of values, what would let minimize the averaged CEP; to verify the determined minimum, testing the performance of 2LP heavily.

The further performance optimization in 2LP (12) with its two optimized parameters in the point (29) or in right-above-starred ones in figure 5 may concern 2LP

$$\mathscr{P}\left(4800, N^*_{\text{HLN}}, 26; 0.2, r^*; R, F, Q_{\text{pass}}\right) \tag{31}$$

with its averaged CEP $p_{\text{EC}}\left(R, F, Q_{\text{pass}}\right)$ to be minimized on integer parallelepiped of values of the ternary $\left\{R, F, Q_{\text{pass}}\right\}$. It's worth to mind that any spotted minimum may move as the function is extremized on other variables. Nevertheless, the ternary minimum for 2LP (31) here could make CEP lower, and the couple $\left\{r^*, N^*_{\text{HLN}}\right\}$ wouldn't have been nonoptimal then due to quasioptimal ranges of PTSDR and HLNN. Note that in this case the optimized $\mathscr{P}_0\left(4800, N^*_{\text{HLN}}, 26\right)$ can be trained with PTSDR r^* so, that gains in further minimization of the performance of 2LP (31) may appear too insignificant.

REFERENCES

1. Arulampalam, G., Bouzerdoum, A. (2003), "A generalized feedforward neural network architecture for classification and regression", *Neural Networks*, vol. 16, iss. 5–6, pp. 561–568.

2. Castillo, P.A., Merelo, J.J., Arenas, M.G., Romero, G. (2007), "Comparing evolutionary hybrid systems for design and optimization of multilayer perceptron structure along training parameters", *Information Sciences*, vol. 177, iss. 14, pp. 2884–2905.

3. Romanuke, V.V. (2014), "Pixel-to-turn standard deviations ratio optimization for training two-layer perceptron on pixel-distorted turned 60-by-80-images in turned objects classification problem", *Visnik Zaporizkogo natsionalnogo universitetu: zbirnik naukovih statey. Fiziko-matematichnI nauki. Matematichne modelyuvannya i prikladna mehanika*, no. 1, pp. 147–170.

4. Romanuke, V.V. (2015), "Two-layer perceptron for classifying scaled-turned-shifted objects by 26 classes general totality of monochrome 60-by-80-images via training with pixel-distorted scaled-turned-shifted images", *Information processing systems*, iss. 7 (132), pp. 98–107.

5. Tsoi, A.C., Back, A. (1995), "Static and dynamic preprocessing methods in neural networks", *Engineering Applications of Artificial Intelligence*, vol. 8, iss. 6, pp. 633–642.

6. Nawi, N.M., Atomi, W.H., Rehman, M.Z. (2013), "The effect of data pre-processing on optimized training of artificial neural networks", *Procedia Technology*, vol. 11, pp. 32–39.

7. Anastassiou, G.A. (2011), "Multivariate sigmoidal neural network approximation", *Neural Networks*, vol. 24, iss. 4, pp. 378–386.

8. Romanuke, V.V. (2015), "Optimizing parameters of the two-layer perceptrons' boosting ensemble training for accuracy improvement in wear state discontinuous tracking model regarding statistical data inaccuracies and shifts", *Problems of tribology*, no. 1, pp. 65–68.

9. Moller, M.F. (1993), "A scaled conjugate gradient algorithm for fast supervised learning", *Neural Networks*, vol. 6, iss. 4, pp. 525–533.

10. Kathirvalavakumar, T., Jeyaseeli Subavathi, S. (2009), "Neighborhood based modified backpropagation algorithm using adaptive learning parameters for training feedforward neural networks", *Neurocomputing*, vol. 72, iss. 16–18, pp. 3915–3921.

ОПТИМАЛЬНОЕ ЧИСЛО НЕЙРОНОВ СКРЫТОГО СЛОЯ ДВУХСЛОЙНОГО ПЕРСЕПТРОНА И СООТНОШЕНИЕ СРЕДНЕКВАДРАТИЧНЫХ ОТКЛОНЕНИЙ ДЛЯ ЕГО ОБУЧЕНИЯ НА ПОВЁРНУТЫХ 60×80-ИЗОБРАЖЕНИЯХ С ПИКСЕЛЬНЫМИ ИСКАЖЕНИЯМИ В КЛАССИФИКАЦИИ ОБЪЕКТОВ С ПОВОРОТАМИ

В. В. Романюк

Хмельницкий национальный университет

ул. Институтская, 11, г. Хмельницкий, 29000, Украина. E-mail: romanukevadimv@mail.ru

Оптимизируется двухслойный персептрон по его двум параметрах для задачи классификации объектов с поворотами. Первым параметром является число нейронов в скрытом слое. Вторым параметром является соотношение среднеквадратичных отклонений пиксельных искажений и поворотов при формировании обучающего множества. Найдено оптимальное число нейронов в скрытом слое двухслойного персептрона вместе с соотношением среднеквадратичных отклонений пиксельных искажений и поворотов для обучения на повёрнутых 60×80-изображениях с пиксельными искажениями. Пара таких оптимальных параметров, впрочем, не является единственной, и приводятся другие варианты оптимальных пар. Двухслойный персептрон при этих параметрах обучается в оптимальном режиме. Его производительность на повёрнутых объектах с 4800 признаками является почти наилучшей. Средний процент ошибок классификации уменьшается до 0,13 %. При максимальной интенсивности искажений поворотами процент ошибок классификации уменьшается до 0,85 %.

Ключевые слова: искажение поворотами, классификация, персептрон, обучение, продуктивность, процент ошибок классификации, оптимизация.

Стаття надійшла 30.10.2015.