V.V. Romanuke*

Polish Naval Academy, Gdynia, Poland

FLEXIBLE SOLUTION OF A 2-LAYER PERCEPTRON OPTIMIZATION BY ITS SIZE AND TRAINING SET SMOOTH DISTORTION RATIO FOR CLASSIFYING SIMPLE-STRUCTURED OBJECTS

Background. Two-layer perceptrons are preferred to complex neural network classifiers when objects to be classified have a simple structure. However, even images consisting of a few geometrically primitive elements on a monotonous background are classified poorly with two-layer perceptron if they are badly distorted (shifted, skewed, and scaled). Performance of two-layer perceptron can be bettered much with modifying its training. This is done by deliberately importing distortions like shifting, skewness, and scaling into a training set, but only controlling volumes of the distortions with a special ratio. Besides, the performance is improved with optimally sizing the hidden layer.

Objective. The goal is to optimize the two-layer perceptron by its size and the ratio for classifying simple-structured objects.

Methods. The objects are monochrome images of enlarged English alphabet capital letters (the EEACL26 dataset) of a medium size 60-by-80. EEACL26 is an infinite artificial dataset, so mathematical models of distorted images are given. Then two-layer perceptrons having various sizes and training set smooth distortion ratios are trained and tested. The performance is evaluated via ultimate-distortion classification error percentage.

Results. Based on statistical evaluations of classification error percentage at ultimate distortions, it is revealed that, while the best ratio should be between 0.01 and 0.02, and an optimal number of neurons in the hidden layer should be between 361 and 390. Sizes closer to 375 are counted as statistically more reliable, whereas the ratios are selected uniformly. Such solution is flexible allowing not only further-training with changing the hidden layer size and ratio, but also a smart initial training for the first passes. Nevertheless, even after the first 100 passes, the two-layer perceptron further-trained for extra 1190 passes by 10 times increasing distortion smoothness performs at 8.91 % of errors at ultimate distortions, which is about 45 % better than a previously known result. At the expected practicable distortions, which are far less, the error rate is 0.8 % that is quite tolerable. But here the main benefit of the optimized two-layer perceptron is its fast operation speed, rather than accuracy.

Conclusions. The obtained flexible solution fits other datasets similar to EEACL26. Number of classes can vary between 20 and 30, and number of features can vary between a few hundred and a few thousand. The stated example of achieving high-performance classification with two-layer perceptrons is a part of the common technique of statistical optimization relating to neural network classifiers. For a more sophisticated dataset of objects, this technique is built and executed in a similar way.

Keywords: classification; shifted-skewed-scaled objects; 2-layer perceptron size; 2-layer perceptron configuration; training set; MATLAB training function; 2-layer perceptron performance.

Introduction

Computer vision controllers and monitors, functioning in real-time currents, work on objects of a finite set of $C \in \mathbb{N} \setminus \{1\}$ classes, which differ significantly from the pattern objects (POs) [1]. Such difference between an N-dimensional tracked object and N-dimensional PO by $N \in \mathbb{N}$ is stated in distinctions in their features, presented as N-dimensional real-valued matrices (RVMs) of a format

$$\mathscr{L} = \underset{d=1}{\overset{N}{\times}} L_d,$$

where the object has L_d features in its d-th dimension, $L_d \in \mathbb{N} \; \forall d = \overline{1, N}$ [2, 3]. For $N \in \{1, 2, 3\}$ the difference can be seen, whereas for $N > 3$ the difference is visualized only in the object plane or space-solid projections [4, 5]. And when one watches the difference, there may be seen how one PO is shifted in relation to its center (towards each of its dimensions), skewed (rotated at a plane or space-solid angle), and scaled. These three most observable key properties have been cited in consecution of catching their view with the human eye [1, 2, 4, 6, 7]. Also there occur more complicated properties of the object, being compared to a pattern one [1, 5, 7]. For instance, it can have a changed number of its

---
* corresponding author: romanukevadimv@gmail.com

dimensions or features [1, 3, 8]. But this rare complication will not be devoted attention to.

Problem statement

For cases of simple-structured objects on a monotonous background (this is about images), it is useful to mathematically consider an object at the classifier's input as a distorted pattern object (DPO), and so there is no difference anymore, but a shifted-skewed-scaled (SSS or S3) object [1, 2]. Classification of S3 objects is hard because of either slow classifiers based on hierarchical multilayered neural networks (neocognitrons [3, 4], cresceptrons [5, 6], convoluteonal neural networks [7, 8]) or poor-performing classifiers based on perceptrons [2, 9, 10]. However, the poor performance of perceptrons over S3 objects is explained just with fails in training them, whereas a hierarchical multilayered neural network is a program structure of a great difficulty and huge memory consumption that makes impossible to accelerate its action. Henceforward, there is a belief and goal to make training a perceptron on S3 objects effective, so it could classify S3 objects with its proper promptitude.

Training set for 2-layer perceptron to classify S3 objects

Theoretically, 2-layer perceptron (2LP) is a universal approximator that fits almost everything if it was identified correctly [2, 10, 11]. 2LP initialization is

$$\mathscr{P}_0(F, H, C; f_{\text{HLTF}}, f_{\text{OLTF}}) \qquad (1)$$

by $F = \prod_{d=1}^{N} L_d$ and a 2LP hidden layer size H in neurons with a hidden layer transfer function f_{HLTF} and an output layer transfer function f_{OLTF}. 2LP (1) is a raw mapping, having

$$F \cdot H + H + H \cdot C + C = H \cdot (F + C + 1) + C \quad (2)$$

non-adjusted weight and bias values. They are adjusted during the three-staged training process, when 2LP (1) input is fed with training sets [9]. On the first stage, 2LP is trained on C pattern F-featured objects, presented in the form of N-dimensional RVMs $\{\mathbf{B}_c = [b_J^{\langle c \rangle}]_{\mathscr{Z}}\}_{c=1}^{C}$ with a subscript J. On the second stage, 2LP input is fed with the training set

$$\{\{\bar{\mathbf{B}}\}_{d=1}^{R}, \{\bar{\mathbf{A}}(p)\}_{p=1}^{P}\} \qquad (3)$$

of $R \in \mathbb{N}$ replicas $\bar{\mathbf{B}} = [b_{wc}]_{F \times C}$ and $P \in \mathbb{N}$ portions of $F \times C$ matrices $\{\bar{\mathbf{A}}(p)\}_{p=1}^{P}$ of $P \cdot C$ DPOs, where c-th column of matrix $\bar{\mathbf{B}}$ corresponds to the c-th class PO, whose matrix $\mathbf{B}_c$ has been reshaped into $F \times 1$ matrix-column, and c-th column of matrix $\bar{\mathbf{A}}(p)$ corresponds to the c-th class DPO. Set (3) is passed through 2LP for $Q \in \mathbb{N}$ times. On the third stage, 2LP is re-trained on the single replica $\bar{\mathbf{B}}$. Clearly, the key stage is the second stage. In a case of S3 objects, this stage might be close to effectiveness if roughly $C < 10$ and $F < 10$, else training on (3) mostly does not converge [1, 12, 13]. But it was shown in articles [2, 9] for $C = 26$ and $F = 4800$ that the problem of swiftly classifying S3 objects could be solved with 2LP, trained with a modified training set

$$\{\{\bar{\mathbf{B}}\}_{d=1}^{R}, \{\tilde{\bar{\mathbf{A}}}(p)\}_{p=1}^{P}\} \text{ by } \tilde{\bar{\mathbf{A}}}(p) = \bar{\mathbf{A}}(p) + \lambda(p) \cdot \mathbf{N}$$

$$\text{at } \lambda(p) = \frac{\lambda_{\max}}{P} \cdot p \text{ for } p = \overline{1, \, P} \qquad (4)$$

instead of (3), where $\lambda(p)$ is a standard deviation (SD) by 4800×26 matrix $\mathbf{N}$ of values of normal variate with zero expectation and unit variance (NVZEUV). POs were modeled with 26 monochrome 60×80 images (M6080Is) $\{\mathbf{B}_c = [b_{uv}^{\langle c \rangle}]_{60 \times 80}\}_{c=1}^{26}$ of enlarged English alphabet capital letters (EEACLs) [2, 7, 9, 10], and, instead of training on S3 M6080Is, there was training on S3 M6080Is with normally distributed pixel distortion (S3 M6080I-NDPD). Also 2LP (1)

$$\mathscr{P}_0(4800, 300, 26; \, \mathscr{G}, \, \mathscr{G}) \qquad (5)$$

with a logarithmic sigmoid transfer function (LSTF) "logsig" $\mathscr{G}$ was attached to integers $R = 2$ and $P = 8$ for the training set in (4), fed the input of 2LP (5) for $Q = 220$ times, and maximal SDs $\eta_{\max} = 0.2$, $\mu_{\max} = 0.2$, $\beta_{\max} = 1$, defining ultimate intensity of scaling, skewing and shifting, respectively [9]. Although maximal SD $\lambda_{\max} = \lambda(8) = 0.08$ for addition in (4) was assigned scientifically, integer $H = 300$ was preset for 2LP (5) empirically. Notwithstanding the groundless $\lambda(8)$ along with 2LP size H, it performed at the ultimate-distortion classification error percentage (CEP) 12.92 % (after the best 2LP had been further-trained, setting 10 times greater P), that is pretty fine for S3 object classifier (under the expected practicable distortions, which are less, the corresponding CEP was 1.64 %, so only one object of 64 ones was misclassified). Nevertheless, CEP for 2LP

classifier over S3 objects as S3 M6080Is might be decreased more if to optimize 2LP in its size H and $\lambda(8)$ simultaneously.

For building a swift classifier of S3 M6080Is on the base of 2LP by optimizing the 2LP size H and value $\lambda(8)$, there are six items to be completed:

1. To configure 2LP for training on S3 M6080I-NDPD.

2. To state mathematical models of S3 M6080Is and S3 M6080I-NDPD.

3. To test the trained 2LP for evaluating its performance as a function of SD $\lambda(8)$ and 2LP size H.

4. To optimize those two factors for identifying 2LP.

5. To train further the identified 2LP until its performance becomes unimprovable (the performance should be higher than performance of 2LP with CEP 12.92 % in article [9]).

6. To circumscribe bounds of S3 objects, which can be classified with the identified optimally-sized 2LP (for optimality in its size and SD for pixel distortion).

All computational processes are coming to be run within MATLAB environment [2, 9, 10], because of MATLAB possesses any needful means to build and simulate 2LP, regulating its parameters, properties or attributes far beyond operatively than being coded and run in C++ or C#. Also MATLAB since R2012a supports powerful enhancements for parallel and cluster computations, which shall be used for training and multi-batch testings to make multiple 2LP performance evaluations.

2LP configuration for training on S3 M6080I-NDPD

Before training on S3 M6080I-NDPD, 2LP

$$\mathscr{P}_0(4800,\ H,\ 26;\ \mathscr{S},\ \mathscr{S}) \qquad (6)$$

is initialized with MATLAB function "feedforwardnet" from Neural Network Toolbox [9]. Any other initialization function (of any other appropriate environment) can be surely used as well. 2LP (6) size H for S3 4800-featured objects with C = 26 must be of order of hundreds or higher. From previous experience, $H \in \{[100;\ 460] \cap \mathbb{N}\}$. 2LP (6) is adapted with weight and bias learning rules by Neural Network Toolbox adapt function "adaptwb". It is going to be trained with MATLAB backpropagation training function "traingda", which will update $4827 \cdot H + 26$. weight and bias values (2) of 2LP (6) according to gradient descent with the adaptive learning rate [2, 14, 15]. Usefulness of 2LP (6) during training will be measured

with its performance function "mse" according to the sum of squared errors. Finally, having preset the minimum performance gradient to 10^{-6}, let the number of epochs be 15,000 in order to prevent long-dragging convergence and to shorten duration for each of Q passes.

Models of S3 M6080Is and S3 M6080I-NDPD

Let the process of forming S3 M6080Is be divided into three phases. These phases come in their succession of completing the previous phase with the next one: scaling, skewing, shifting. This is actually reverse to consecution of catching S3 objects' view with the human eye.

During the first phase $\mathbf{B}_c$ is scaled with the scaling map σ into the c-th class EEACL as $V \times M$ matrix $\mathbf{Z}_c(p)$ by SD $\eta(p)$ and the scale coefficient $\varsigma(\eta(p))$, where

$$\sigma[\mathbf{B}_c,\ \varsigma(\eta(p))] = \mathbf{Z}_c(p),\ \ \eta(p) = \frac{\eta_{\max}}{P} \cdot p,$$
$$\varsigma(\eta(p)) = \eta(p)\xi(p) + 1 \ \text{at}\ \ p = \overline{1,\ P} \qquad (7)$$

with the value $\xi(p)$ of NVZEUV $\Xi(p)$, which is reraffled until $\varsigma(\eta(p)) > 0$. In MATLAB the map σ in (7) is supported with MATLAB function "imresize" [2, 9, 16]. M6080I $\mathbf{B}_c$ is enlarged by $\varsigma(\eta(p))$ times through (7) if $\varsigma(\eta(p)) > 1$, and $\mathbf{B}_c$ is reduced by $\dfrac{1}{\varsigma(\eta(p))}$ times through (7) if $\varsigma(\eta(p)) < 1$, and $\mathbf{B}_c$ remains itself if $\varsigma(\eta(p)) = 1$.

During the second phase $\mathbf{Z}_c(p)$ is skewed at an angle $\rho(p)$ in degrees by SD $\mu(p)$ to $V \times M$ matrix $\mathbf{T}_c(p)$, where

$$\mathbf{T}_c(p) = 1 - \tau[1 - \mathbf{Z}_c(p),\ \rho(p)],\ \ \mu(p) = \frac{\mu_{\max}}{P} \cdot p,$$
$$\rho(p) = \frac{180}{\pi} \cdot \mu(p)\theta(p)\ \text{at}\ \ p = \overline{1,\ P} \qquad (8)$$

with the value $\theta(p)$ of NVZEUV $\Theta(p)$. The map τ in (8), supported with MATLAB function "imrotate", skews the input negative $1 - \mathbf{Z}_c(p)$ in counterclockwise direction if $\rho(p) > 0$, and for $\rho(p) < 0$ it is turned clockwise, while for $\rho(p) = 0$ the matrix $\mathbf{Z}_c(p)$ remains itself. Before shifting in the third phase, $\mathbf{T}_c(p)$ is re-formatted back into 60×80 matrix $\mathbf{R}_c(p)$, still representing the c-th class EEACL. If $\varsigma(\eta(p)) > 1$ then lines with indices I and columns with indices J in $\mathbf{T}_c(p)$ are discarded, where

$$I = \{\{\overline{1,\ N_V}\},\ \{\overline{61 + N_V,\ V}\}\},$$

$$J = \{\{\overline{1,\ N_M}\},\ \{\overline{81 + N_M,\ M}\}\},$$

$$N_V = \Omega\left(\frac{V}{2}\right) - 30 + \left(\frac{1 + \operatorname{sign}\psi_V}{2} \cdot \operatorname{sign}|\psi_V|\right)$$

$$\times \operatorname{sign}\left[\frac{V}{2} - \Omega\left(\frac{V}{2}\right)\right],$$

$$N_M = \Omega\left(\frac{M}{2}\right) - 40 + \left(\frac{1 + \operatorname{sign}\psi_M}{2} \cdot \operatorname{sign}|\psi_M|\right)$$

$$\times \operatorname{sign}\left[\frac{M}{2} - \Omega\left(\frac{M}{2}\right)\right],$$

and $\Omega(\alpha)$ is a function, returning the integer part of the number α, calculated by values $\{\psi_V,\ \psi_M\}$ of two independent NVZEUVs. These NVZEUVs are raffled every time, when the function $\Omega(\alpha)$ is applied. If $\varsigma(\eta(p)) < 1$ then $\mathbf{T}_c(p)$ is padded from left for N_{left} columns of ones and from right for N_{right} columns of ones, and it is padded from top for N_{top} lines of ones and from bottom for N_{bottom} lines of ones for contouring the skewed-scaled image $\mathbf{T}_c(p)$ with the background white color [2, 9, 16], where

$$N_{\text{left}} = \Omega\left(\frac{80 - M}{2}\right) + \left(\frac{1 + \operatorname{sign}\psi_M}{2} \cdot \operatorname{sign}|\psi_M|\right)$$

$$\times \operatorname{sign}\left[\frac{M}{2} - \Omega\left(\frac{M}{2}\right)\right],$$

$$N_{\text{right}} = 80 - M - N_{\text{left}},$$

$$N_{\text{top}} = \Omega\left(\frac{60 - V}{2}\right) + \left(\frac{1 + \operatorname{sign}\psi_V}{2} \cdot \operatorname{sign}|\psi_V|\right)$$

$$\times \operatorname{sign}\left[\frac{V}{2} - \Omega\left(\frac{V}{2}\right)\right],$$

$$N_{\text{bottom}} = 60 - V - N_{\text{top}}.$$

The c-th class representative $\mathbf{R}_c(p)$ is shifted in two subphases [9, 17]. By SD

$$\beta(p) = \frac{\beta_{\max}}{P} \cdot p \quad \text{for} \quad p = \overline{1,\ P} \tag{9}$$

it is shifted for $X[\beta(p)]$ pixels horizontally and for $Y[\beta(p)]$ pixels vertically, where

$$X[\beta(p)] = \varphi(8\beta(p) \cdot \gamma_1(p))$$

$$\times \frac{1 - \operatorname{sign}(|\varphi(8\beta(p) \cdot \gamma_1(p))| - 80)}{2}$$

$$+ 80 \cdot \frac{1 + \operatorname{sign}(|\varphi(8\beta(p) \cdot \gamma_1(p))| - 80)}{2},$$

$$Y[\beta(p)] = \varphi(6\beta(p) \cdot \gamma_2(p))$$

$$\times \frac{1 - \operatorname{sign}(|\varphi(6\beta(p) \cdot \gamma_2(p))| - 60)}{2}$$

$$+ 60 \cdot \frac{1 + \operatorname{sign}(|\varphi(6\beta(p) \cdot \gamma_2(p))| - 60)}{2},$$

and function $\varphi(\alpha)$ rounds α to the nearest integer less than or equal to α with values $\{\gamma_1(p),\ \gamma_2(p)\}$ of two independent NVZEUVs. During the first shift subphase, after horizontal shift and before vertical one, the matrix $\mathbf{R}_c(p) = [r_{uv}^{\langle c \rangle}(p)]_{60 \times 80}$ is changed into matrix $\mathbf{S}_c(p) = [s_{uv}^{\langle c \rangle}(p)]_{60 \times 80}$, where

$$s_{uv}^{\langle c \rangle}(p) = 1 \quad \text{for} \quad v = \overline{1,\ X[\beta(p)]}$$

$$\text{and} \quad s_{uv}^{\langle c \rangle}(p) = r_{ut}^{\langle c \rangle}(p) \quad \text{at} \quad t = v - X[\beta(p)]$$

$$\text{for} \quad v = \overline{X[\beta(p)] + 1, 80}$$

$$\forall u = \overline{1, 60} \quad \text{by} \quad X[\beta(p)] > 0,$$

$$s_{uv}^{\langle c \rangle}(p) = r_{ut}^{\langle c \rangle}(p) \quad \text{at} \quad t = v - X[\beta(p)]$$

$$\text{for} \quad v = \overline{1, 80 + X[\beta(p)]}$$

$$\text{and} \quad s_{uv}^{\langle c \rangle}(p) = 1 \quad \text{for} \quad v = \overline{81 + X[\beta(p)], 80}$$

$$\forall u = \overline{1, 60} \quad \text{by} \quad X[\beta(p)] < 0,$$

$$s_{uv}^{\langle c \rangle}(p) = r_{uv}^{\langle c \rangle}(p) \quad \forall u = \overline{1, 60}$$

$$\text{and} \quad \forall v = \overline{1, 80} \quad \text{by} \quad X[\beta(p)] = 0.$$

The second subphase in shifting the plane M6080I is the vertical shifting, where matrix $\mathbf{S}_c(p)$ is changed into matrix $\mathbf{A}_c(p) = [a_{uv}^{\langle c \rangle}(p)]_{60 \times 80}$:

$$a_{uv}^{\langle c \rangle}(p) = s_{rv}^{\langle c \rangle}(p) \quad \text{at} \quad r = u + Y[\beta(p)]$$

$$\text{for} \quad u = \overline{1, 60 - Y[\beta(p)]}$$

$$\text{and} \quad a_{uv}^{\langle c \rangle}(p) = 1 \quad \text{for} \quad u = \overline{61 - Y[\beta(p)], 60}$$

$$\forall v = \overline{1, 80} \quad \text{by} \quad Y[\beta(p)] > 0,$$

$$a_{uv}^{\langle c \rangle}(p) = 1 \quad \text{for} \quad u = \overline{1,\ -Y[\beta(p)]}$$

$$\text{and} \quad a_{uv}^{\langle c \rangle}(p) = s_{rv}^{\langle c \rangle}(p) \quad \text{at} \quad r = u + Y[\beta(p)]$$

$$\text{for} \quad u = \overline{-Y[\beta(p)] + 1, 60} \quad \forall v = \overline{1, 80}$$

$$\text{by} \quad Y[\beta(p)] < 0$$

$$a_{uv}^{\langle c \rangle}(p) = s_{uv}^{\langle c \rangle}(p) \quad \forall u = \overline{1, 60}$$

$$\text{and} \quad \forall v = \overline{1, 80} \quad \text{by} \quad Y[\beta(p)] = 0.$$

Hereon the third phase of forming S3 M6080I $\mathbf{A}_c(p)$ is complete for $p = \overline{1, P}$.

Subsequently the p-th full portion of 60×80 matrices $\{\mathbf{A}_c(p)\}_{c=1}^{26}$ is reshaped into 4800×26 matrix $\overline{\mathbf{A}}(p)$ for $p = \overline{1, P}$. Then p-th portion of the full batch of S3 M6080I-NDPD for the modified training set is formed due to (4). Now 2LP (6) by fixing its size $H \in \{[100; 460] \cap \mathbb{N}\}$ is ready for training under parameters

$$\eta_{max} = 0.2, \ \mu_{max} = 0.2, \ \beta_{max} = 1, \ R = 2,$$
$$P = 8, \ Q = 100 \tag{10}$$

and some SD $\lambda(8)$. Once the best 2LP under parameters (10) is obtained, it will be further-trained until its performance becomes unimprovable [9].

Evaluating the performance of the trained 2LP

For presetting some SD $\lambda(8)$ it is useful to denote the pixel-to-shift-skew-scale standard deviations ratio (PS3SDR) r_{PS3SDR}, that shall define how much of pixel distortion should be imported into matrices $\{\overline{\mathbf{A}}(p)\}_{p=1}^P$. PS3SDR can be assigned as one of the statements:

$$\frac{\lambda_{max}}{\eta_{max}}, \ \frac{\lambda_{max}}{\mu_{max}}, \ \frac{\lambda_{max}}{\beta_{max}}.$$

And for definiteness, let PS3SDR be assigned via SD of the hardest by its influence distortion, especially since $\beta_{max} = 1$:

$$r_{PS3SDR} = \frac{\lambda_{max}}{\beta_{max}} = \lambda_{max}. \tag{11}$$

From previous experience, $\lambda_{max} \in [0.01; 0.1]$ and so $r_{PS3SDR} \in [0.01; 0.1]$. However, value $r_{PS3SDR} = 0$ will be tested for comparisons.

Mainly the performance of the trained 2LP (6), whose post-training configuration is denoted by

$$\mathscr{P}(4800, H, 26; \mathfrak{S}, \mathfrak{S}; \eta_{max},$$
$$\mu_{max}, \beta_{max}, r_{PS3SDR}, R, P, Q)$$
$$= \mathscr{P}(4800, H, 26; \mathfrak{S}, \mathfrak{S}; 0.2,$$
$$0.2, 1, r_{PS3SDR}, 2, 8, 100) \tag{12}$$

is CEP

$$p_{max}(H, r_{PS3SDR}) = \frac{q_{max}(H, r_{PS3SDR})}{b \cdot C} \cdot 100 \tag{13}$$

with a classification errors number (CEN) $q_{max}(H, r_{PS3SDR})$ by b full batches of C S3 objects, feeding the input of 2LP (12). Therefore, for optimization of 2LP classifier of S3 M6080Is, there is an optimization problem

$$[H^* \ r_{PS3SDR}^*]$$
$$\in \arg\left(\min_{[H \ r_{PS3SDR}] \in \{[100; 460] \cap \mathbb{N}\} \times [0; 0.1]} \{p_{max}(H, r_{PS3SDR})\}\right) \tag{14}$$

on a line rectangle

$$\{[100; 460] \cap \mathbb{N}\} \times [0; 0.1] \tag{15}$$

of 2LP (12) sizes and PS3SDR (11) values. While 2LP (12) is tested, its input is fed with $b = 400$ full batches of S3 M6080Is, formed by maximal SDs

$$\eta = \eta_{max} = 0.2, \ \mu = \mu_{max} = 0.2, \ \beta = \beta_{max} = 1. \tag{16}$$

We do not use SDs in their ranges like

$$\eta \in [0; \eta_{max}] = [0; 0.2], \ \mu \in [0; \mu_{max}] = [0; 0.2],$$
$$\beta \in [0; \beta_{max}] = [0; 1],$$

because, using the previous experience, there no special SDs that could induce discontinuities of CEP plotted against H and r_{PS3SDR}.

It is clear that the segment of PS3SDR (11) $[0; 0.1]$ must be sampled. A successive sweep of 2LP (12) on the segment of PS3SDR $[0; 0.1]$ may be fulfilled on the pointed line rectangle (15)

$$\{100 + 20k\}_{k=0}^{18} \times \{0.01i\}_{i=0}^{10}$$
$$\subset \{[100; 460] \cap \mathbb{N}\} \times [0; 0.1]. \tag{17}$$

The purpose of sweeping 2LP (12) through 209 points on the point rectangle (17) is to narrow domain (15), where the problem (14) solution would be enclosed. So, after the first approximation, an evaluation of surface (13) on (17) is in Fig. 1 (the CEP-against-H view is intentionally overshadowed).

The evaluation of surface (13) in Fig. 1 is a mean of 10 2LPs in each of those 209 points. It is obvious that Fig. 1 (namely, the CEP-against-H view) allows to narrow problem (14) to a problem

$$[H^* \ r_{PS3SDR}^*]$$
$$\in \arg\left(\min_{[H \ r_{PS3SDR}] \in \{[360; 460] \cap \mathbb{N}\} \times [0; 0.1]} \{p_{max}(H, r_{PS3SDR})\}\right). \tag{18}$$

Second approximation shall either show the minimum point of surface (13) or narrow the problem (18) towards it.
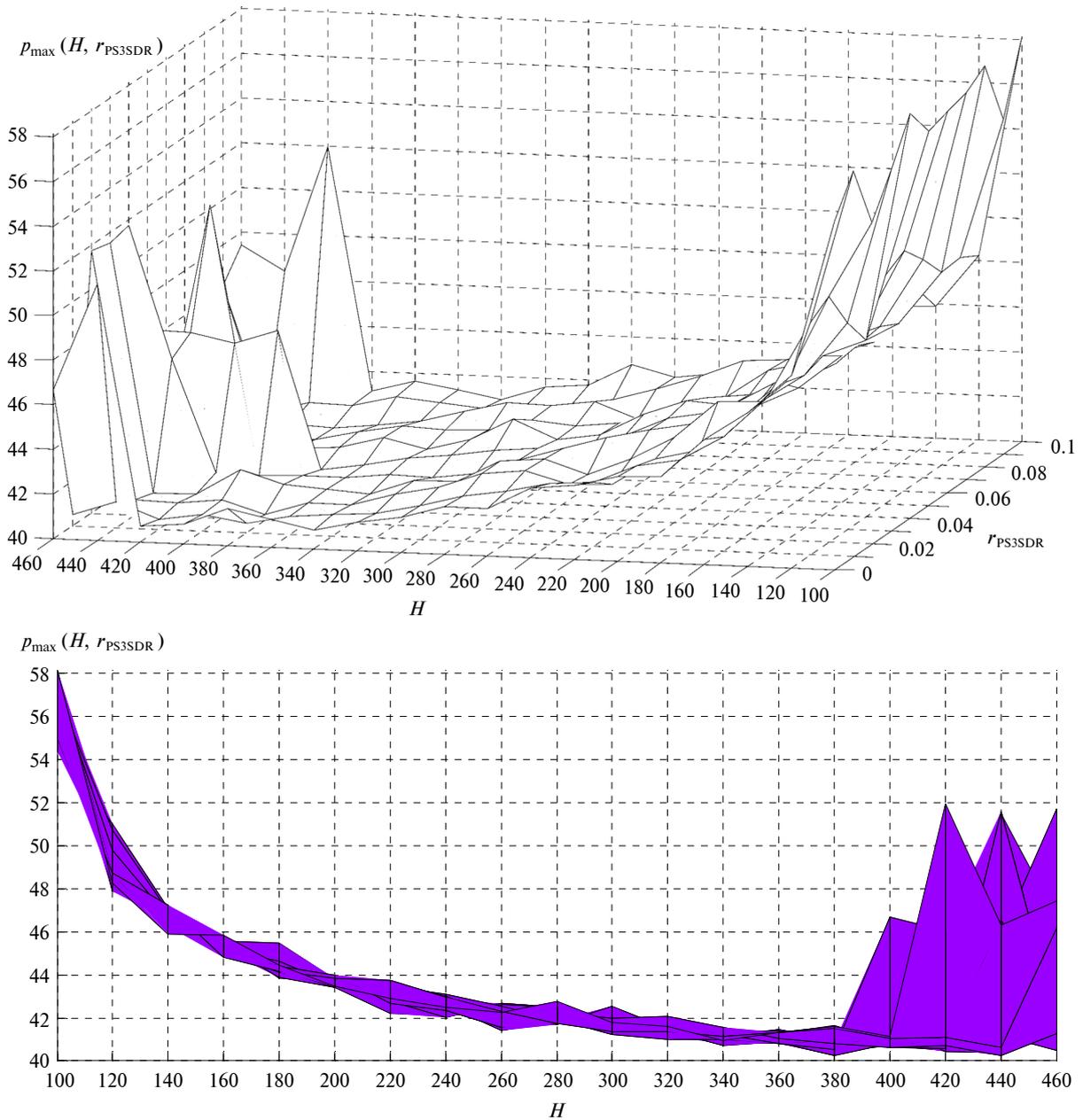
Fig. 1. An evaluation of surface (13) for 2LP (12) on the point rectangle (17) with the CEP-against-H view beneath the surface

Solution of problem (14)

A re-evaluation of surface (13) on a pointed domain

$$\{360 + 20k\}_{k=0}^{5} \times \{0.01i\}_{i=0}^{10}$$
$$\subset \{[360; 460] \cap \mathbb{N}\} \times [0; 0.1] \qquad (19)$$

under minimum in (18) is shown in Fig. 2 (the CEP-against-H view and the CEP-against-PS3SDR view are intentionally overshadowed). The CEP-against-H view allows to conclude that $H^* \in \{[361; 390] \cap \mathbb{N}\}$.

The CEP-against-PS3SDR view addresses the PS3SDR value 0.02 to us, whereas the PS3SDR value 0.08 might have been ignored because a seeming minimum at $H = 440$ and $r_{PS3SDR} = 0.08$ is unconvincing (see Fig. 3). Remember that $r_{PS3SDR} = 0.08$ was a solution of the similar problem in [9], where 2LP had been optimized by only PS3SDR. Here, nonetheless, $r_{PS3SDR} = 0.08$ is off because a mean of CEPs over a sub-domain

$$\{360, 380, 400\} \times \{0.01i\}_{i=0}^{10} \qquad (20)$$
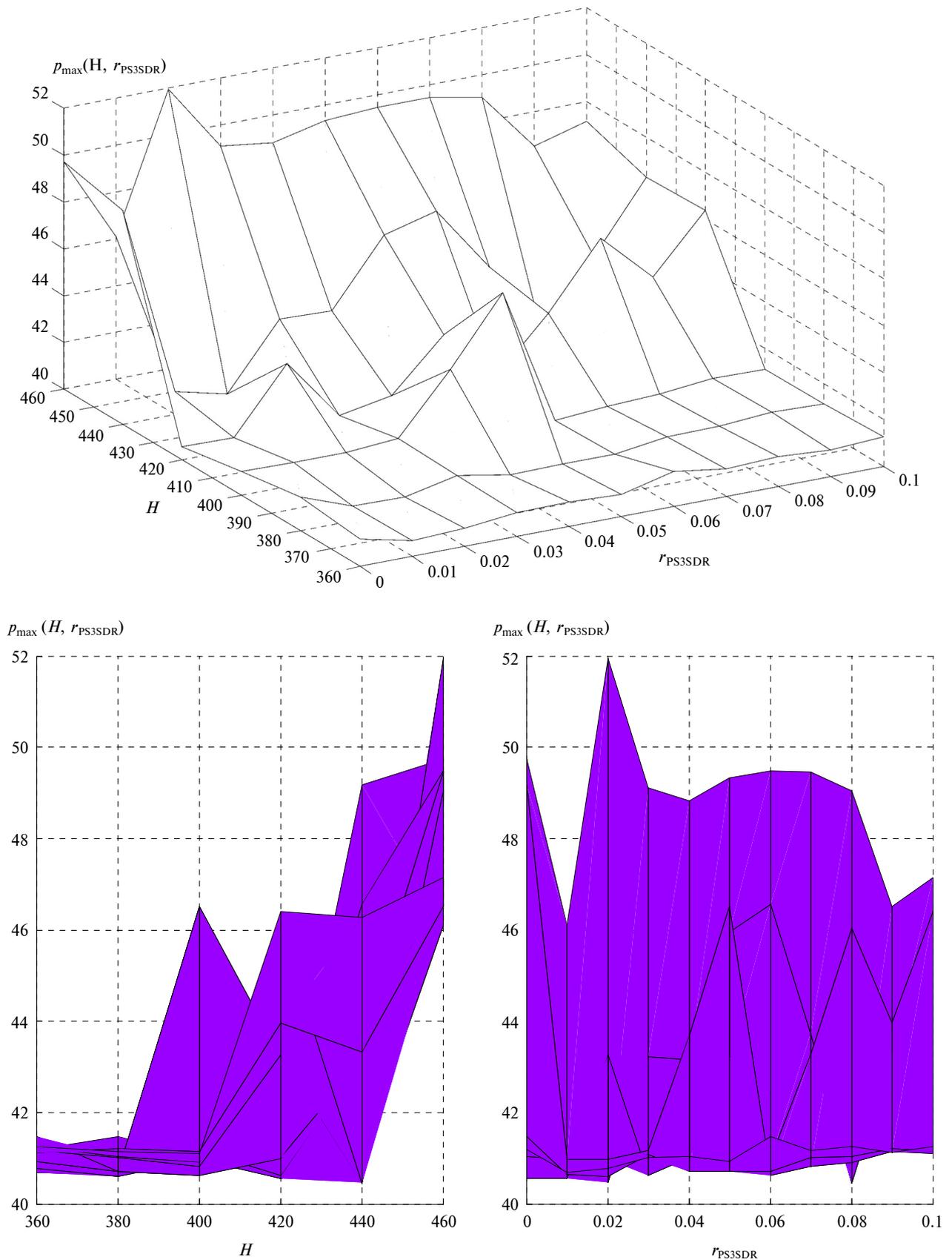
Fig. 2. A re-evaluation of surface (13) for 2LP (12) on the point rectangle (19) with the CEP-against-H and CEP-against-PS3SDR views beneath the surface, where the re-evaluation is a mean of 20 2LPs in each of those 66 points (10 2LPs relating to every point in Fig. 1 are used as well)
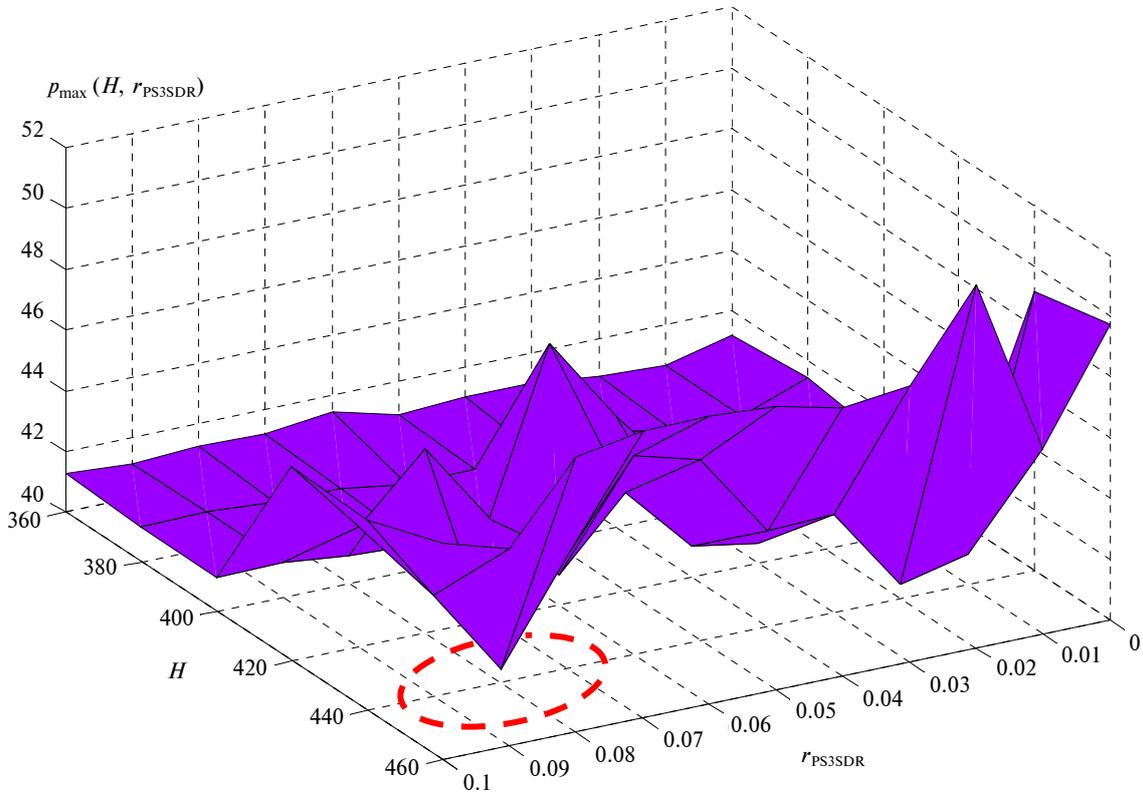
Fig. 3. Vicinity of the point $[H \; r_{\text{PS3SDR}}] = [440 \; 0.08]$ encloses only points, at which CEP is too high, so the small CEP at $H = 440$ and $r_{\text{PS3SDR}} = 0.08$ is pure randomness
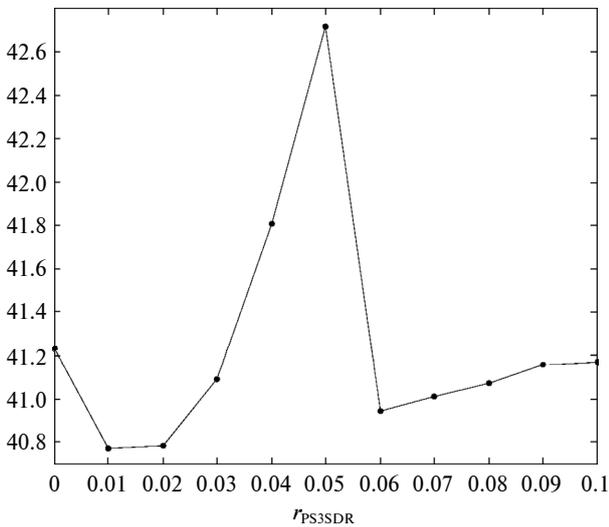
Fig. 4. A mean of 60 CEPs (from 20 2LPs times 3 sizes of 2LP) over sub-domain (20), where minima at the PS3SDR values 0.01 and 0.02 are close
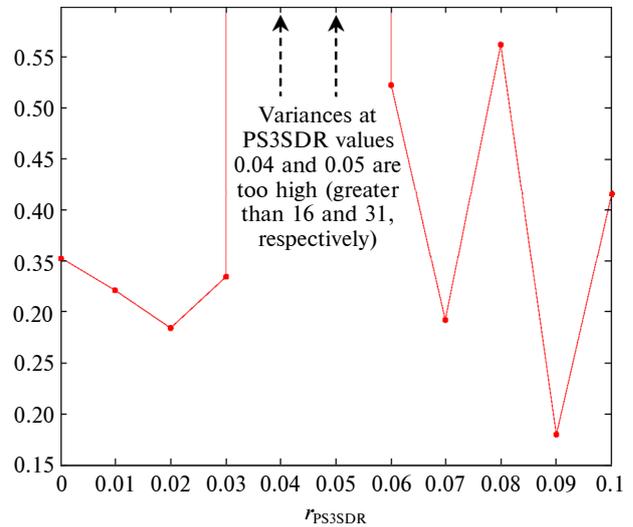
Fig. 5. CEP variance of 60 2LPs (the PS3SDR values 0.04 and 0.05 drop out as they thrice produced CEPs greater than 96 %, i. e. three 2LPs failed)

taken against PS3SDRs is worse for $r_{\text{PS3SDR}} > 0.03$ (see this in Fig. 4). Moreover, CEPs for $r_{\text{PS3SDR}} > 0.03$ become much unstable (see CEP variance of 60 2LPs against PS3SDRs in Fig. 5).

Another reason to accept $r^{*}_{\text{PS3SDR}} = 0.02$ is that this is very stable PS3SDR, producing small CEPs (Fig. 6). On the other hand, setting $r_{\text{PS3SDR}} = 0.02$ gives a lesser number of 2LPs with high CEPs (Fig. 7).
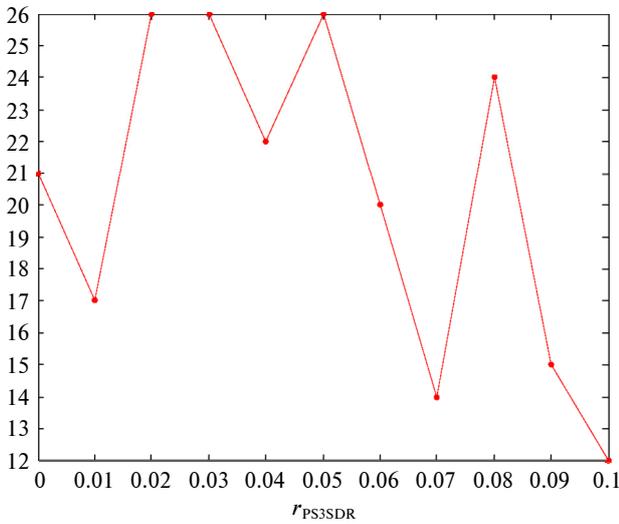
Fig. 6. Number of 2LPs whose CEP is less than 40 %

Fig. 7. Number of 2LPs whose CEP is higher than 41 %

As total number of 2LPs (for each point of PS3SDR) is 120 here (only those six 2LP sizes from Fig. 2 are considered), then a score of $r^*_{\text{PS3SDR}} = 0.02$ in Fig. 7 is about 16 % and 11 % better than scores of $r_{\text{PS3SDR}} = 0.03$ and $r_{\text{PS3SDR}} = 0.05$, respectively.

Apparently, no single point can be asserted as a solution of problem (14). However, before training 2LP

$$\mathscr{P}(4800, H^*, 26; \mathscr{S}, \mathscr{S}; 0.2, 0.2, 1, r^*_{\text{PS3SDR}}, 2, 8, 100)$$

further, a single 2LP size and the PS3SDR value must be set. A too great size makes training longer. Therefore, $r^*_{\text{PS3SDR}} = 0.02$ and $H^* = 370$.

Further-training

Firstly, a 2LP

$$\mathscr{P}(4800, 370, 26; \mathscr{S}, \mathscr{S}; 0.2, 0.2, 1, 0.02, 2, 8, 100) \quad (21)$$

is obtained. For training 2LP (21) further, let P be increased to 80. Training will be stopped if performance improvement fails 50 times in succession. Performance of 2LP (21) is $p_{\max}(370, 0.02) \approx 39.46$, which is nearly the same as CEPs in Fig. 2 for $H = 360$ and $H = 380$. Having further-trained 2LP (21) for extra 1190 passes (denote this 2LP by 1190FT-2LP) by increasing smoothness of distortions with $P = 80$ (see in article [18], how training is smoothed for improving the 2LP performance), performance exceeds that 12.92 % at ultimate distortions (16): now CEP is 8.91 %, which is about 45 % better. The 1190FT-2LP performs very well at lesser distortions.

Its CEP across the increasing distortions tied to shifting denoted by $p^*(\beta; 370, 0.02)$ is seen in Fig. 8. Being compared to analogous CEP in [9] denoted by $p^*_{0.08}(\beta)$, its gain is obvious. Distortions by SDs (16) appear ultimate, so the expected practicable SDs are less. They are about up to $\beta = 0.7$ by the corresponding $\eta = 0.14$ and $\mu = 0.14$, where CEP is 0.8 % that is quite tolerable.

Difference between successfully classified (the left side images) and misclassified S3 M6080Is by maximal SDs in (16) is shown in Fig. 9, where six level of distortions (connected to the shifting SDs) are shown. The misclassified images appear so much distorted, that the classifier cannot properly "see" them [9]. Seemingly, bad shifting appearing the most intense distortion influences worst on the classifier. Scaling influences much as well (but neither scaling nor skewing causes so much misclassification as shifting does). Even under twice lesser SDs, the misclassified EEACLs appear very much distorted (the right side images).

Although the 1190FT-2LP could be further-trained more, with probably greater P, increment of P has its limit, after which the performance is unimprovable [9, 18]. Another possibility is slightly varying PS3SDR before every batch of extra passes is done. Changing the 2LP size is possible as well:

1) when H is decreased, the respective indices (which are randomly generated) in both layers and in biases to the hidden layer are deleted;

2) when H is increased, a new indices are randomly inserted into both layers and into biases to the hidden layer (the previous indices are offset),
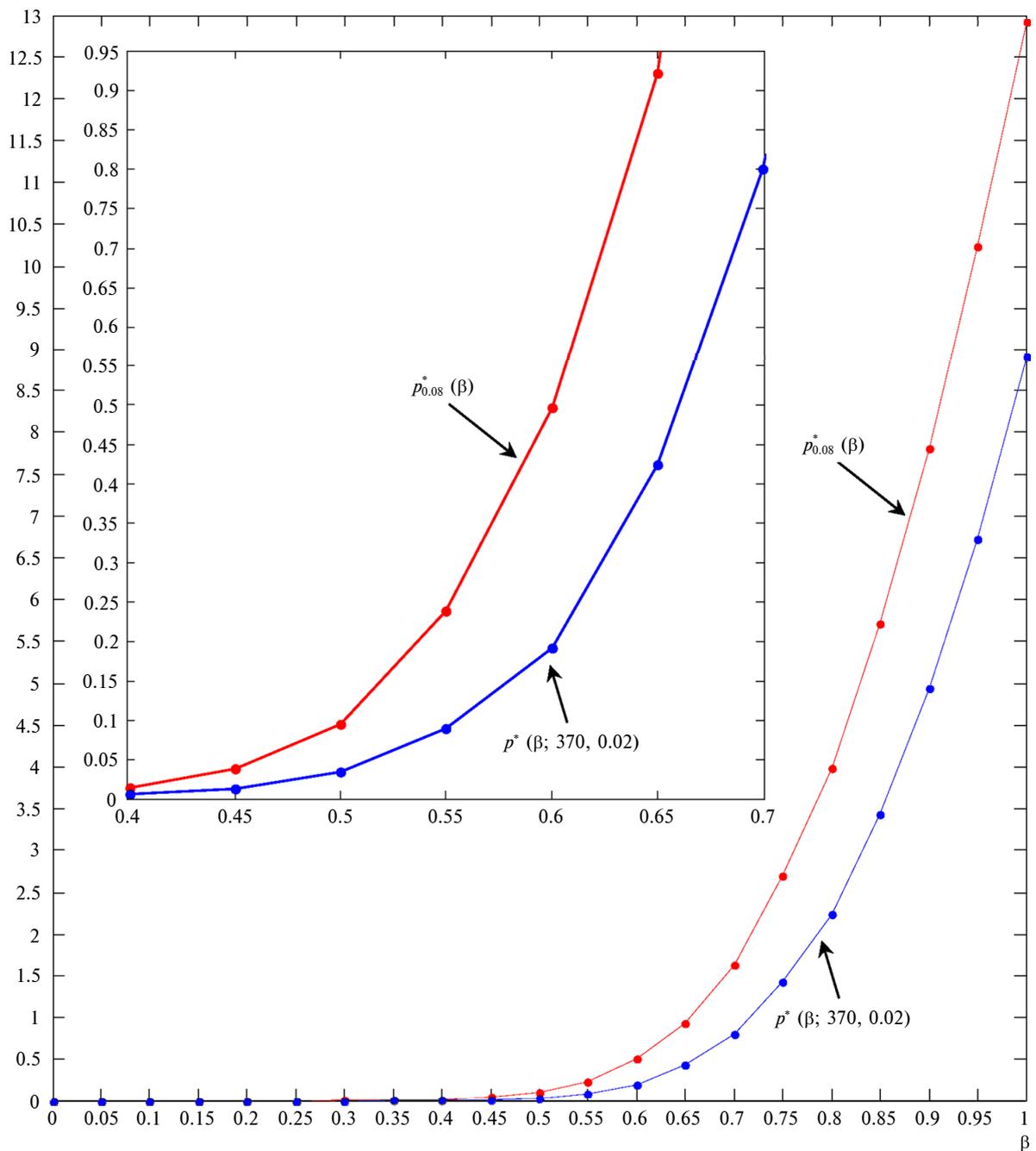
Fig. 8. A disclosure of CEP across the increasing distortions tied to shifting

and values for their elements are randomly selected from a normal distribution (the expectation and variance should be close to those in the corresponding layer).

Thus, the solution of problem (14) can be re-represented wider as

$$[H^* \; r^*_{\text{PS3SDR}}] \in \{[361;390] \cap \mathbb{N}\} \times [0.01;0.02] \quad (22)$$

by referring to Figs. 4 and 5 along with Figs. 6 and 7. This is a rough decision, but a researcher can freely change the hidden layer size from 361 to 390 neurons by varying PS3SDR between 0.01 and 0.02, and performances for 2LP (21) and further-trained 2LPs are expected to be the best. Such a line rectangle solution is more reliable and flexible than a solution with a single 2LP size and a single PS3SDR value [19, 20].

Fig. 9. EEACLs misclassified (the right side images) and classified successfully (the left side images) by 1190FT-2LP

Propagation of solution (22)

Solution (22) has been obtained for an infinite dataset of EEACLs [7, 9, 10, 16]. If the dataset is changed to a similar one, however, solution (22) is expected to keep its optimality for 2LPs (see it in Fig. 10, whereon statistically more reliable points are lined out thicker; nonetheless, thickness along a ho-

rizontal line remains the same because varying those PS3SDRs influences less on CEP). The similarity consists in the following items:

1. Number of classes is roughly between 20 and 30. A case with $C = 26$ is perfect.

2. Number of features is roughly between a few hundred and a few thousand. A case with $F = 4800$ or about that is perfect.

3. POs and DPOs (S3 POs) have a monotonous background, and their structure is of either a few geometrically primitive elements (for images) or a few common specificities (for objects having a one dimension or four and more). Color images (of objects having three dimensions) should not have gradient colors.
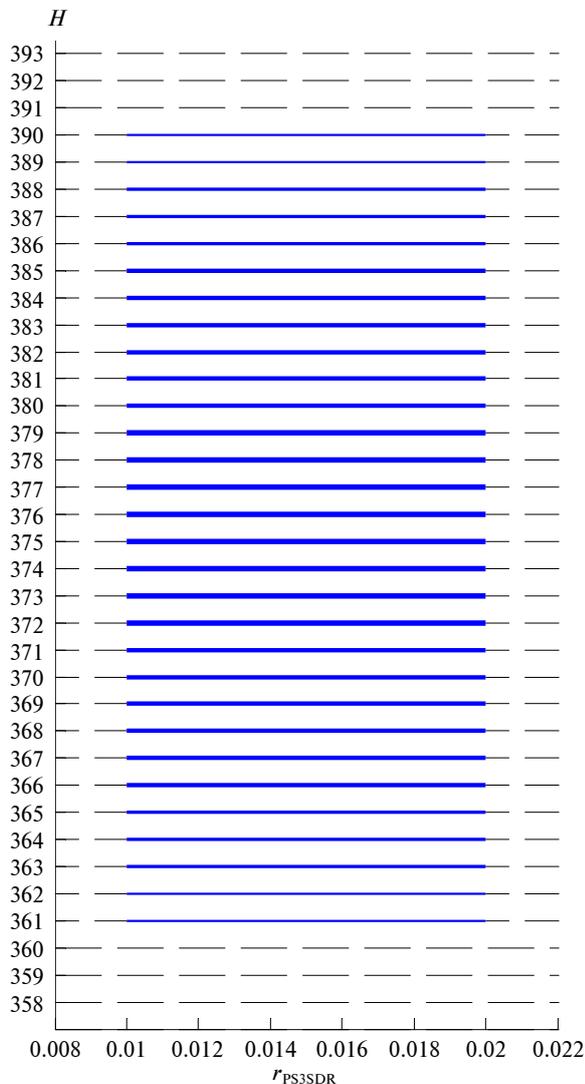
Fig. 10. Solution (22) with points lined out thicker as statistically more reliable

The stated example of achieving high-performance classification of S3 POs with 2LPs is a part of the common technique of statistical optimization relating to neural network classifiers [21]. For a more sophisticated dataset of objects, this technique is built and executed in a similar way, but 2LPs would not be effective there anymore, so more complicated neural networks are optimized on a lot of other parameters (so-called hyperparameters). The main requirement for treating the solution (22) applicable is similarity between a dataset to be classified and the EEACL26 dataset. At that, a slight modification of the training process at its second stage may be done. Instead of linearly increasing SDs by (4), (7), (8), and (9), they may be taken in an increasing order, so that

$$\lambda(p) < \lambda(p+1), \ \ \eta(p) < \eta(p+1),$$
$$\mu(p) < \mu(p+1), \qquad (23)$$
$$\beta(p) < \beta(p+1) \ \ \forall \ p = \overline{1, P-1}.$$

Moreover, it is not necessary to create sequences (23) simultaneously. Such diversity shall help in preventing overfitting.

Conclusions

Flexibility of solution (22) allows not only further-training with changing the 2LP size and PS3SDR, but also a smart initial training for the first Q passes. An impact of such training is expected to be. This, however, is not going to expand efficient application of 2LPs — the efficiency exists just for datasets of simple-structured objects (remember those three items above).

Here, the main benefit of the optimized 2LP is its fast operation speed, rather than accuracy. Compared to slower classifiers built on more complicated and accurate neural networks, 2LPs like (21) or that 1190FT-2LP can process a few times more objects in a second, without losing much accuracy. So, optimized 2LPs can be used in combination with complex classifiers. Nevertheless, 2LPs (along with other simple classifiers like learning vector quantization networks, support vector machines, radial basis networks, probabilistic neural networks, generalized regression neural networks) remain the best for classifying objects of single dimension ("column" data). But, what is crucial, 2LPs having just two uncertain parameters (its size by H and PS3SDR or similar training set smooth distortion ratio) are optimized for a definite classification problem in the easiest way.

List of literature

1. *Haykin S.O.* Neural Networks and Learning Machines. – Upper Saddle River, NJ: Pearson, 2009. – 936 p.
2. *Romanuke V.V.* 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 // Inform. Process. Systems. – 2015. – Iss. 7 (132). – P. 98−107.
3. *Cardoso B., Wichert A.* Noise tolerance in a Neocognitron-like network // Neural Networks. – 2014. – **49**. – P. 32−38.
4. *Fukushima K.* Training multi-layered neural network neocognitron // Neural Networks. – 2013. – **40**. – P. 18−31.
5. *Weng J., Ahuja N., Huang T.S.* Learning recognition and segmentation using the Cresceptron // Int. J. Comp. Vision. – 1997. – **25**, iss. 2. – P. 109−143.
6. *Weng J., Ahuja N., Huang T.S.* Cresceptron: A self-organizing neural network which grows adaptively // Proc. Int. Joint Conf. Neural Networks, Baltimore, Maryland. – 1992. – **I**. – P. 576−581.
7. *Romanuke V.V.* Appropriate number and allocation of ReLUs in convolutional neural networks // Naukovi Visti NTUU KPI. – 2017. – № 1. – P. 69−78.
8. *Learning* pooling for convolutional neural network / M. Sun, Z. Song, X. Jiang et al. // Neurocomputing. – 2017. – **224**. – P. 96−104.
9. *Romanuke V.V.* Classifying scaled-turned-shifted objects with optimal pixel-to-scale-turn-shift standard deviations ratio in training 2-layer perceptron on scaled-turned-shifted 4800-featured objects under normally distributed feature distortion // Electrical Control Commun. Eng. – 2017. – **13**. – P. 45−54.
10. *Romanuke V.V.* A 2-layer perceptron performance improvement in classifying 26 turned monochrome 60-by-80-images via training with pixel-distorted turned images // Naukovi Visti NTUU KPI. – 2014. – № 5. – P. 55−62.
11. *Baum E.B.* On the capabilities of multilayer perceptrons // J. Complexity. – 1988. – **4**, iss. 3. – P. 193−215.
12. *An efficient* hidden layer training method for the multilayer perceptron / C. Yu, M.T. Manry, J. Li, P.L. Narasimha // Neurocomputing. – 2006. – **70**, iss. 1-3. – P. 525−535.
13. *Bhattacharya U., Parui S.K.* On the rate of convergence of perceptron learning // Pattern Recognition Lett. – 1995. – **16**, iss. 5. – P. 491−497.
14. *Moller M.F.* A scaled conjugate gradient algorithm for fast supervised learning // Neural Networks. – 1993. – **6**, iss. 4. – P. 525−533.
15. *On-line* neural training algorithm with sliding mode control and adaptive learning rate / A. Nied, S.I.Jr. Seleme, G.G. Parma, B.R. Menezes // Neurocomputing. – 2007. – **70**, iss. 16-18. – P. 2687−2691.
16. *Romanuke V.V.* Two-layer perceptron for classifying flat scaled-turned-shifted objects by additional feature distortions in training // J. Uncertain Syst. – 2015. – **9**, № 4. – P. 286−305.
17. *Romanuke V.V.* An attempt for 2-layer perceptron high performance in classifying shifted monochrome 60-by-80-images via training with pixel-distorted shifted images on the pattern of 26 alphabet letters // Radio Electronics, Computer Science, Control. – 2013. – № 2. – P. 112−118.
18. *Romanuke V.V.* Dependence of performance of feed-forward neuronet with single hidden layer of neurons against its training smoothness on noised replicas of pattern alphabet // Herald of Khmelnytskyi National University. Tech. Sci. – 2013. – № 1. – P. 201−206.
19. *Interval* optimization combined with point estimate method for stochastic security-constrained unit commitment / M. Zhou, S. Xia, G. Li, X. Han // Int. J. Electrical Power & Energy Systems. – 2014. – **63**. – P. 276−284.
20. *Optimal* unit sizing for small-scale integrated energy systems using multi-objective interval optimization and evidential reasoning approach / F. Wei, Q.H. Wu, Z.X. Jing et al. // Energy. – 2016. – **111**. – P. 933−946.
21. *Romanuke V.V.* Optimal training parameters and hidden layer neuron number of two-layer perceptron for generalised scaled object classification problem // Inform. Technol. Management Sci. – 2015. – **18**. – P. 42−48.

References

[1] S.O. Haykin, *Neural Networks and Learning Machines.* Upper Saddle River, NJ: Pearson, 2009.
[2] V.V. Romanuke, "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", *Inform. Process. Syst.*, iss. 7, pp. 98−107, 2015.
[3] B. Cardoso and A. Wichert, "Noise tolerance in a Neocognitron-like network", *Neural Networks*, vol. 49, pp. 32−38, 2014. doi: 10.1016/j.neunet.2013.09.007
[4] K. Fukushima, "Training multi-layered neural network neocognitron", *Neural Networks*, vol. 40, pp. 18−31, 2013. doi: 10.1016/j.neunet.2013.01.001

[5] J. Weng *et al.*, "Learning recognition and segmentation using the Cresceptron", *Int. J. Comp. Vision*, vol. 25, iss. 2, pp. 109−143, 1997.

[6] J. Weng *et al.*, "Cresceptron: A self-organizing neural network which grows adaptively", in *Proc. Int. Joint Conf. Neural Networks*, Baltimore, Maryland, 1992, vol. I, pp. 576−581. doi: 10.1109/IJCNN.1992.287150

[7] V.V. Romanuke, "Appropriate number and allocation of ReLUs in convolutional neural networks", *Naukovi Visti NTUU KPI*, no. 1, pp. 69−78, 2017. doi: 10.20535/1810-0546.2017.1.88156

[8] M. Sun *et al.*, "Learning pooling for convolutional neural network", *Neurocomputing*, vol. 224, pp. 96−104, 2017. doi: 10.1016/j.neucom.2016.10.049

[9] V.V. Romanuke, "Classifying scaled-turned-shifted objects with optimal pixel-to-scale-turn-shift standard deviations ratio in training 2-layer perceptron on scaled-turned-shifted 4800-featured objects under normally distributed feature distortion", *Electrical Control Commun. Eng.*, vol. 13, pp. 45−54, 2017. doi: 10.1515/ecce-2017-0007

[10] V.V. Romanuke, "A 2-layer perceptron performance improvement in classifying 26 turned monochrome 60-by-80-images via training with pixel-distorted turned images", *Naukovi Visti NTUU KPI*, no. 5, pp. 55−62, 2014.

[11] E.B. Baum, "On the capabilities of multilayer perceptrons", *J. Complexity*, vol. 4, iss. 3, pp. 193−215, 1988. doi: 10.1016/0885-064X(88)90020-9

[12] C. Yu *et al.*, "An efficient hidden layer training method for the multilayer perceptron", *Neurocomputing*, vol. 70, iss. 1-3, pp. 525−535, 2006. doi: 10.1016/j.neucom.2005.11.008

[13] U. Bhattacharya and S.K. Parui, "On the rate of convergence of perceptron learning", *Pattern Recognition Lett.*, vol. 16, iss. 5, pp. 491−497, 1995. doi: 10.1016/0167-8655(95)00124-Y

[14] M.F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning", *Neural Networks*, vol. 6, iss. 4, pp. 525−533, 1993. doi: 10.1016/S0893-6080(05)80056-5

[15] A. Nied *et al.*, "On-line neural training algorithm with sliding mode control and adaptive learning rate", *Neurocomputing*, vol. 70, iss. 16-18, pp. 2687−2691, 2007. doi: 10.1016/j.neucom.2006.07.019

[16] V.V. Romanuke, "Two-layer perceptron for classifying flat scaled-turned-shifted objects by additional feature distortions in training", *J. Uncertain Syst.*, vol. 9, no. 4, pp. 286−305, 2015.

[17] V.V. Romanuke, "An attempt for 2-layer perceptron high performance in classifying shifted monochrome 60-by-80-images via training with pixel-distorted shifted images on the pattern of 26 alphabet letters", *Radio Electronics, Computer Science, Control*, no. 2, pp. 112−118, 2013. doi: 10.15588/1607-3274-2013-2-18

[18] V.V. Romanuke, "Dependence of performance of feed-forward neuronet with single hidden layer of neurons against its training smoothness on noised replicas of pattern alphabet", *Herald of Khmelnytskyi National University. Tech. Sci.*, no. 1, pp. 201−206, 2013.

[19] M. Zhou *et al.*, "Interval optimization combined with point estimate method for stochastic security-constrained unit commitment", *Int. J. Electrical Power & Energy Systems*, vol. 63, pp. 276−284, 2014. doi: 10.1016/j.ijepes.2014.06.012

[20] F. Wei *et al.*, "Optimal unit sizing for small-scale integrated energy systems using multi-objective interval optimization and evidential reasoning approach", *Energy*, vol. 111, pp. 933−946, 2016. doi: 10.1016/j.energy.2016.05.046

[21] V.V. Romanuke, "Optimal training parameters and hidden layer neuron number of two-layer perceptron for generalised scaled object classification problem", *Inform. Technol. Management Sci.*, vol. 18, pp. 42−48, 2015. doi: 10.1515/itms-2015-0007

В.В. Романюк

ГНУЧКИЙ РОЗВ'ЯЗОК ЗАДАЧІ ОПТИМІЗАЦІЇ ДВОШАРОВОГО ПЕРСЕПТРОНА ЗА ЙОГО РОЗМІРОМ І СПІВВІДНОШЕННЯМ ГЛАДКИХ СПОТВОРЕНЬ У НАВЧАЛЬНІЙ МНОЖИНІ ДЛЯ КЛАСИФІКАЦІЇ ОБ'ЄКТІВ ПРОСТОЇ СТРУКТУРИ

Проблематика. Двошаровим персептронам надається перевага над складними нейронними мережами, коли об'єкти для класифікації мають просту структуру. Однак навіть зображення, що складаються з декількох геометрично примітивних елементів на монотонному фоні, класифікуються двошаровим персептроном погано, якщо вони сильно спотворені (зсунуті, перекошені та масштабовані). Продуктивність двошарового персептрона може бути значно покращена модифікацією його навчання. Це робиться навмисним внесенням спотворень, подібних зсуву, перекошеності та масштабуванню, у навчальну множину, але тільки за допомогою контролю об'ємів цих спотворень спеціальним співвідношенням. Крім того, продуктивність покращується підбором оптимального розміру прихованого шару.

Мета дослідження. Метою роботи є оптимізація двошарового персептрона за його розміром і згаданого співвідношення для класифікації об'єктів простої структури.

Методика реалізації. Об'єктами є монохромні зображення збільшених великих літер англійського алфавіту (набір даних EEACL26) середнього розміру 60×80. EEACL26 є нескінченним штучним набором даних, тому даються математичні моделі спотворених зображень. Далі навчаються і тестуються двошарові персептрони, що мають різні розміри та співвідношення гладких спотворень у навчальній множині. Продуктивність оцінюється за відсотком помилок класифікації на граничних спотвореннях.

Результати дослідження. Виходячи зі статистичних оцінок відсотка помилок класифікації на граничних спотвореннях, виявлено, що найкраще співвідношення повинно бути між 0,01 та 0,02, а оптимальне число нейронів у прихованому шарі має бути між 361 та 390. Розміри ближче до 375 вважаються статистично більш надійними, тоді як співвідношення вибираються рів-

номірно. Такий розв'язок є гнучким, дозволяючи не тільки подальше навчання зі зміною розміру прихованого шару та співвідношення, а й витончене першопочаткове навчання на перших проходах. Проте навіть після перших 100 проходів двошаровий персептрон, навчаний далі на 1190 екстрапроходах зі збільшеною в 10 разів гладкістю спотворень, працює на рівні 8,91 % помилок на граничних спотвореннях, що приблизно на 45 % краще, ніж попередній відомий результат. На рівні практично очікуваних спотворень, які значно менші, рівень помилок становить 0,8 %, що цілком прийнятно. Однак тут головною вигодою оптимізованого двошарового персептрона є не точність, а його швидкодія.

Висновки. Отриманий гнучкий розв'язок підходить іншим наборам даних, подібних EEACL26. Число класів може варіюватися між 20 і 30, а число ознак – між декількома сотнями і декількома тисячами. Наведений приклад досягнення високопродуктивної класифікації двошаровими персептронами є частиною загальної методики статистичної оптимізації стосовно нейромережевих класифікаторів. Для більш складних наборів даних об'єктів ця методика будується і виконується у подібному стилі.

Ключові слова: класифікація; зсунуті масштабовані скошені об'єкти; розмір двошарового персептрона; конфігурація двошарового персептрона; навчальна множина; Matlab-функція для навчання; продуктивність двошарового персептрона.

В.В. Романюк

ГИБКОЕ РЕШЕНИЕ ЗАДАЧИ ОПТИМИЗАЦИИ ДВУХСЛОЙНОГО ПЕРСЕПТРОНА ПО ЕГО РАЗМЕРУ И СООТНОШЕНИЮ ГЛАДКИХ ИСКАЖЕНИЙ В ОБУЧАЮЩЕМ МНОЖЕСТВЕ ДЛЯ КЛАССИФИКАЦИИ ОБЪЕКТОВ ПРОСТОЙ СТРУКТУРЫ

Проблематика. Двухслойные персептроны предпочитаемы сложным нейронным сетям, когда объекты для классификации имеют простую структуру. Однако даже изображения, состоящие из нескольких геометрически примитивных элементов на монотонном фоне, классифицируются двухслойным персептроном плохо, если они сильно искажены (сдвинуты, перекошены и отмасштабированы). Производительность двухслойного персептрона может быть значительно улучшена модификацией его обучения. Это делается преднамеренным внесением искажений, подобных сдвигу, перекошенности и масштабированию, в обучающее множество, но только с помощью контроля объемов этих искажений специальным соотношением. Кроме того, производительность улучшается подбором оптимального размера скрытого слоя.

Цель исследования. Целью работы является оптимизация двухслойного персептрона по его размеру и упомянутого соотношения для классификации объектов простой структуры.

Методика реализации. Объектами являются монохромные изображения увеличенных заглавных букв английского алфавита (набор данных EEACL26) среднего размера 60×80. EEACL26 является бесконечным искусственным набором данных, поэтому даются математические модели искаженных изображений. Затем обучаются и тестируются двухслойные персептроны, имеющие различные размеры и соотношения гладких искажений в обучающем множестве. Производительность оценивается через процент ошибок классификации на предельных искажениях.

Результаты исследования. Исходя из статистических оценок процента ошибок классификации на предельных искажениях, обнаружено, что наилучшее соотношение должно быть между 0,01 и 0,02, а оптимальное число нейронов в скрытом слое должно быть между 361 и 390. Размеры ближе к 375 считаются статистически более надежными, тогда как соотношения выбираются равномерно. Такое решения является гибким, позволяя не только дальнейшее обучение с изменением размера скрытого слоя и соотношения, но также и изящное начальное обучение на первых проходах. Тем не менее даже после первых 100 проходов двухслойный персептрон, обученный далее на 1190 экстрапроходах с увеличенной в 10 раз гладкостью искажений, работает на уровне 8,91 % ошибок при предельных искажениях, что примерно на 45 % лучше, чем предыдущий известный результат. На уровне практически ожидаемых искажений, которые значительно меньше, уровень ошибок составляет 0,8 %, что вполне сносно. Однако здесь главной выгодой оптимизированного двухслойного персептрона является не точность, а его быстродействие.

Выводы. Полученное гибкое решение подходит другим наборам данных, подобным EEACL26. Число классов может варьироваться между 20 и 30, а число признаков – между несколькими сотнями и несколькими тысячами. Приведенный пример достижения высокопроизводительной классификации двухслойными персептронами является частью общей методики статистической оптимизации в отношении нейросетевых классификаторов. Для более сложных наборов данных объектов эта методика строится и исполняется в подобном стиле.

Ключевые слова: классификация; сдвинутые масштабированные скошенные объекты; размер двухслойного персептрона; конфигурация двухслойного персептрона; обучающее множество; Matlab-функция для обучения; производительность двухслойного персептрона.