ISSN 2078–5097. Вісн. Львів. ун-ту. Сер. прикл. матем. та інф. 2017. Вип. 25. С. 112–129
Visnyk of the Lviv University. Series Appl. Math. and Informatics. Issue 25. P. 112–129

UDC 004.032.26::004.93

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

V. Romanuke

Khmelnitskiy National University,
Instytutska Str., 11, Khmelnitskiy, 29016

A classification problem is considered. Objects to be classified are presumed distorted with linear scaling effect. The classifier is two-layer perceptron. The object model is the monochrome 60-by-80-image of the enlarged English alphabet capital letter. Thus general totality is formed of monochrome 60-by-80-images of alphabet letters, and it is of 26 classes. The goal is to show how hidden layer neurons number in the two-layer perceptron and an element of the topological configuration for training it can be optimized for scaled objects classification problem. The topological configuration element is a pixel-to-scale standard deviations ratio. This ratio allows importing feature-distorted objects into the training set. The feature distortion is actually formed by adding the normal noise with zero expectation and variance regulated by the ratio. The two-layer perceptron is modeled, trained and tested within MATLAB. Factually, this is an optimization problem of two variables, where the two-layer perceptron performance is optimized in the sense of decreasing classification error percentage. The percentage is a function evaluated on the Cartesian product of the ranges of hidden layer neurons number and pixel-to-scale standard deviations ratio. This product is a horizontally-striped rectangle which is subsequently sampled to a lattice. The optimization result is 150 neurons in the hidden layer and the ratio equal to 0.02, allowing to identify a classifier which produces an error only on an object from 37 objects whose scaling effect distortion is maximal.

Key words: scaled objects classification, two-layer perceptron, monochrome image, training set, classification error percentage, optimal hidden layer neurons number, optimal pixel-to-scale standard deviations ratio.

1. SCALING IN OBJECTS CLASSIFICATION

Effects of scaling are common for object recognition (classification) systems, working via photographing or scanning. If even there is no scaling in a photograph or a scan, say, faces of a child and an adult are always viewed differently in their outlines, although they are captured at the same distance. Other scaling effects occur when the object data are processed or transformed, bringing linear stretching or reduction of the imaged object. Notwithstanding this, scaling in objects classification mustn't lower the classification quality, what is achieved and maintained with scaling-proof classifiers.

2. WAYS OF IMPROVING THE CLASSIFIER FOR OPTIMIZING ITS PERFORMANCE OVER SCALED OBJECTS

Whatever the scaling-proof classifier is, there at any time may appear a scaling amplitude which breaks its proofness. Thus there will be a need to optimize the classifier. Anyway classification error percentage (CEP) in the classifier performance is non-zero, and such CEP is tried to have it lowered anyhow. Especially, when it is about the classifier performance over scaled objects. This performance depends on the scaling range and its peak values, on the general totality of those objects and number of classes in it. The greater all these properties and attributes are, the harder it is to identify a classifier, producing low CEP, which is tolerable [1, 2]. So, the classifier performance optimization over scaled objects is always actual as the scaling range is extending, the classes' number is increasing or the general totality is becoming denser.

But along with lowering CEP the improvement of the classifier concerns also its operation speed. And here classifiers based on hierarchical multilayered neural networks or convolutional neural networks lose much [3]. The matter is that they are pretty slow-acting, though being capable of robust visual pattern recognition (object classification) through learning. Besides, those neural networks are grown huge in resources consumption as, saying, the scaling range is extended. Multilayered perceptron, quite the contrary, performs fast and easy, but its performance over scaled objects needs an optimization. Main ways of improving the perceptron for classifying the scaled objects are selecting its neurons number optimally and optimal configuration of the training set [1]. However, the optimal neurons number and the training set optimal configuration are determined deeply on the object type and its features, not to mention the number of the general totality classes and others above.

3. GOAL OF ARTICLE AND TASKS

Before stating the goal of this article, the type of objects for their classification must be appointed. Once the object type is modeled, the general totality and number of classes will be defined. Only then the article goal can be stated explicitly.

Generally, the goal is to show how neurons number in the perceptron and an element of the topological configuration for training it can be optimized for scaled objects classification problem. Certainly, that the show is going to be stated over some pattern of those objects. So, its results are locally applied, although the way of optimization of the perceptron performance will be saved out.

May the type of objects for their classification be the monochrome image. It is very convenient as the monochrome image is naturally watched with its any feature distortions. Moreover, the monochrome image is easily modeled for the object, and number of its features is directly calculated on the image format.

For classifying monochrome images it is adequate to have two-layer perceptron (2LP). Perceptrons are described and simulated best in MATLAB, having powerful Neural Network Toolbox. Subsequently, a MATLAB function for training the perceptron ought to be chosen, and a model of a specific training set will be stated. Then boundaries for the neurons number in the perceptron single hidden layer are to be given. Also boundaries for a distinctive parameter of the specific training set are to be evaluated. This parameter is that element of the topological configuration for training, helping in acceleration of the training process over scaled images. Having boundaries for the hidden layer neurons number (HLNN) and boundaries for the training set distinctive parameter, there is the Cartesian product of two ranges, whereupon CEP is to be minimized on that product. The solution of the correspondingly stated math problem will be verified as testing the trained 2LP by those two optimized parameters. Finally, suggestions for further 2LP performance optimization will conclude this article.

4. GENERAL TOTALITY AND NUMBER OF CLASSES BY THE OBJECT MODEL

An appropriate format for the monochrome image is 60×80. This medium format allows accelerating the investigation procedures and acquiring the classification results. The file format is bitmap, which is naturally coded by MATLAB with ones and zeros. Thus 60-by-80-image is modeled as 60×80 matrix of ones and zeros, and there is the finite general totality of ones-and-zeros 60×80 matrices, containing altogether 2^{4800} monochrome images.

May there 26 enlarged English alphabet capital letters [1] be imaged within the general totality. This gives 26 classes in it. Neither the small number, nor the great, what is acceptable for hitting the article goal at the proper term.

For configuring the advanced training on sample sets of scaled images from such general totality, there is the training set distinctive parameter, having involved the pixel distortion [1, 4]. Pixel distortions are added to the linear scaling effect, and these two types of distortion are attributed by their standard deviations (SD) [1]. Hence, the explicit goal of this article is to minimize CEP in the scaled objects classification problem on the pattern of the general totality of ones-and-zeros 60×80 matrices, modeling 2^{4800} monochrome images, where optimal HLNN in 2LP and optimal pixel-to-scale standard deviations ratio (PSSDR) for its training on pixel-distorted scaled 60-by-80-images (PDS6080I) would be applied.

5. MATLAB FUNCTION FOR TRAINING THE PERCEPTRON

There is the backpropagation algorithm, having several specialized methods to train 2LP. The fastest method in backpropagation algorithm for training on medium-number-featured objects is implemented as the training MATLAB function "traingda". This function effectively trains the perceptron, updating weight and bias values according to gradient descent with adaptive learning rate [1]. Henceforward, the training MATLAB function "traingda" will be used in all training processes below. Then, testing the trained 2LP under previously assigned HLNN N_{HLN} and PSSDR r, there will be obtained the function $p_{\mathrm{error}}\left(r,\, N_{\mathrm{HLN}}\right)$ on the Cartesian product of HLNN and PSSDR ranges, whose values are averaged CEP over a distortion type range of SD.

6. MODEL OF PDS6080I

The training set for 2LP to classify scaled 60-by-80-images (S6080I) is formed in $F \in \mathbb{N}$ stages. This training set is filled with PDS6080I, modeled as 4800×26 matrices of PDS6080I $\left\{\mathbf{A}_{\mathrm{PDS6080I}}^{\langle k \rangle}\right\}_{k=1}^{F}$, whose q-th column is the q-th class representative, reshaped into 4800-length-column:

$$\mathbf{A}_{\mathrm{PDS6080I}}^{\langle k \rangle} = \left[\tilde{\tilde{a}}_{jq}\left(k\right)\right]_{4800 \times 26} = \mathbf{A}_{\mathrm{S6080I}}^{\langle k \rangle} + \sigma_{\mathrm{PD}}^{\langle k \rangle} \cdot \mathbf{\Xi} \tag{1}$$

by SD

$$\sigma_{\mathrm{PD}}^{\langle k \rangle} = \frac{k}{F} \cdot \sigma_{\mathrm{PD}}^{\langle \max \rangle} \quad \forall\, k = \overline{1,\, F} \tag{2}$$

and its maximum $\sigma_{\mathrm{PD}}^{\langle \max \rangle} > 0$ at 4800×26 matrix $\mathbf{\Xi}$ of values of normal variate (NV) with zero expectation and unit variance (ZEUV), where $\mathbf{A}_{\mathrm{S6080I}}^{\langle k \rangle}$ is 4800×26 matrix of S6080I, whose q-th column is the same q-th class representative, 4800-length-column-reshaped, $q = \overline{1,\, 26}$. The matrix $\mathbf{A}_{\mathrm{S6080I}}^{\langle k \rangle} = \left[\tilde{a}_{jq}\left(k\right)\right]_{4800 \times 26}$ is formed by concatenating horizontally

4800-length-column-reshaped matrices $\left\{\tilde{\mathbf{A}}_q(k)\right\}_{q=1}^{26}$, where q-th S6080I as the matrix $\tilde{\mathbf{A}}_q(k)=\left[\tilde{a}_{uv}^{\langle q\rangle}(k)\right]_{60\times 80}$ is the q-th class representative. Note, that generally there in left side of (1) is not the ones-and-zeros matrix, though it is included into the training set that will feed the input of 2LP. But as soon as elements of this matrix are compared to the value 0.5 via 0.5-crossing comparator, it already belongs to the general totality of ones-and-zeros 60×80 matrices, modeling 2^{4800} monochrome images. Namely, if $\tilde{\tilde{a}}_{jq}(k)\leqslant 0.5$ then the pixel of vertical-horizontal coordinates $\{j, q\}$ is set to the black color, else this pixel is set to the white (in MATLAB, the white color is coded with ones, and the black is coded with zeros).

For each k-th stage of forming the training set by (1) and (2), every matrix $\tilde{\mathbf{A}}_q(k)$ is obtained after scaling separately the q-th class non-distorted representative as the ones-and-zeros matrix $\mathbf{A}_q=\left(a_{uv}^{\langle q\rangle}\right)_{60\times 80}$. An S6080I can be formed within MATLAB by means of its function “imresize” [1] as the map

$$\tilde{\mathbf{A}}_q(k)=\rho\left(\mathbf{A}_q,\, \varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right),\, 60,\, 80\right) \tag{3}$$

with the scale coefficient $\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)$ by SD

$$\sigma_{\text{scale}}^{\langle k\rangle}=\frac{k}{F}\cdot\sigma_{\text{scale}}^{\langle\max\rangle} \quad \forall\ k=\overline{1,\, F} \tag{4}$$

for $\sigma_{\text{scale}}^{\langle\max\rangle}>0$, $q=\overline{1,\, 26}$. Mark it out that pixel distortion SD (2) and scale SD (4) are increased simultaneously. The scale coefficient

$$\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)=\sigma_{\text{scale}}^{\langle k\rangle}\xi(k)+1 \tag{5}$$

is determined by the value $\xi(k)$ of NV with ZEUV, raffled at the k-th stage for each q separately. If occurs $\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)\leqslant 0$ then the corresponding NV with ZEUV is re-raffled until $\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)>0$. The input image $\mathbf{A}_q$ is enlarged by $\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)$ times within the map (3) if $\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)>1$; the input image $\mathbf{A}_q$ is reduced by $\dfrac{1}{\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)}$ times within the map (3) if $\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)<1$; the input image $\mathbf{A}_q$ remains non-scaled if $\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)=1$ or (5) is rounded to 1 due to that $\left|\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)-1\right|<0.006$, and $\tilde{\mathbf{A}}_q(k)=\mathbf{A}_q$ as corollary. The scaled by $\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)\neq 1$ image is the matrix $\mathbf{S}\left(q,\, \sigma_{\text{scale}}^{\langle k\rangle}\right)$ of the intermediate format $V\times H$. If $\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right)>1$ then the scaled image is cropped in the following way. Integers

$$N_V=\eta\left(\frac{V}{2}\right)-30+\left(\frac{1+\operatorname{sign}\zeta_V}{2}\cdot\operatorname{sign}\left|\zeta_V\right|\right)\cdot\operatorname{sign}\left[\frac{V}{2}-\eta\left(\frac{V}{2}\right)\right], \tag{6}$$

$$N_H = \eta\left(\frac{H}{2}\right) - 40 + \left(\frac{1+\operatorname{sign}\zeta_H}{2}\cdot\operatorname{sign}\left|\zeta_H\right|\right)\cdot\operatorname{sign}\left[\frac{H}{2}-\eta\left(\frac{H}{2}\right)\right], \tag{7}$$

where $\eta(x)$ is a function, returning the integer part of the number x, are calculated by the values $\{\zeta_V, \zeta_H\}$ of two independent NV with ZEUV, raffled every time, when the function $\eta(x)$ is applied. Then in the matrix $\mathbf{S}\left(q, \sigma_{\text{scale}}^{\langle k\rangle}\right)$ lines of their numbers

$$\left\{\left\{\overline{1, N_V}\right\}, \left\{\overline{61+N_V, V}\right\}\right\} \tag{8}$$

and columns of their numbers

$$\left\{\left\{\overline{1, N_H}\right\}, \left\{\overline{81+N_H, H}\right\}\right\} \tag{9}$$

are discarded. If $\varsigma\left(\sigma_{\text{scale}}^{\langle k\rangle}\right) < 1$ then the reduced image is contoured rectangularly with the background white color: the matrix $\mathbf{S}\left(q, \sigma_{\text{scale}}^{\langle k\rangle}\right)$ is padded from left for

$$N_{\text{left}} = \eta\left(\frac{80-H}{2}\right) + \left(\frac{1+\operatorname{sign}\zeta_H}{2}\cdot\operatorname{sign}\left|\zeta_H\right|\right)\cdot\operatorname{sign}\left[\frac{H}{2}-\eta\left(\frac{H}{2}\right)\right] \tag{10}$$

columns of ones and from right for

$$N_{\text{right}} = 80 - H - N_{\text{left}} \tag{11}$$

columns of ones, and it is padded from top for

$$N_{\text{top}} = \eta\left(\frac{60-V}{2}\right) + \left(\frac{1+\operatorname{sign}\zeta_V}{2}\cdot\operatorname{sign}\left|\zeta_V\right|\right)\cdot\operatorname{sign}\left[\frac{V}{2}-\eta\left(\frac{V}{2}\right)\right] \tag{12}$$

lines of ones and from bottom for

$$N_{\text{bottom}} = 60 - V - N_{\text{top}} \tag{13}$$

lines of ones. After that the map (3) finally returns S6080I as 60×80 matrix $\tilde{\mathbf{A}}_q(k)$ into the set $\left\{\tilde{\mathbf{A}}_q(k)\right\}_{q=1}^{26}$ for completing the addition (1).

7. BOUNDARIES FOR HLNN

Obviously, boundaries $N_{\text{HLN}}^{\langle\min\rangle}$ and $N_{\text{HLN}}^{\langle\max\rangle}$ for the segment $\left[N_{\text{HLN}}^{\langle\min\rangle}; N_{\text{HLN}}^{\langle\max\rangle}\right]$ of HLNN N_{HLN} must be given so that 2LP at any $N_{\text{HLN}} \in \left[N_{\text{HLN}}^{\langle\min\rangle}; N_{\text{HLN}}^{\langle\max\rangle}\right]\bigcap\mathbb{N}$ would be trained appropriately. Influence of traintime duration counts much less than CEP, and HLNN N_{HLN} is accepted if 2LP just can be trained. This “can” means the training process is not of hang-like, and overtraining won’t occur. Taking into account the number 4800 of the image features and number 26 of classes, let the left valid value of HLNN be 100. At that HLNN cannot exceed 350 as by $N_{\text{HLN}} > 350$ 2LP is no more being trained yet on non-distorted representatives. Therefore the boundaries for HLNN are $N_{\text{HLN}}^{\langle\min\rangle} = 100$ and $N_{\text{HLN}}^{\langle\max\rangle} = 350$, and the segment of HLNN is $[100; 350]$. Later, for each of 251 values from this segment, the averaged CEP value at some PSSDR will be determined to obtain the function $p_{\text{error}}\left(r,\, N_{\text{HLN}}\right)$ value.

8. BOUNDARIES FOR PSSDR

From pixel distortion SD (2) and scale SD (4) it follows that [1] PSSDR

$$r = \frac{\sigma_{\text{PD}}^{\langle\max\rangle}}{\sigma_{\text{scale}}^{\langle\max\rangle}}\,. \tag{14}$$

Boundaries $r_{\min}$ and $r_{\max}$ for the segment $\left[r_{\min};\, r_{\max}\right]$ of PSSDR r ought to be evaluated such that after the training process 2LP would maintain its capability to classify both S6080I and PDS6080I. Classification of S6080I is basic. By the way, PSSDR for training and PSSDR in PDS6080I for testing the PDS6080I-trained 2LP are different principally. The range of (14) is not profoundly influenced with the image features number or the number of classes. However, it influences on the traintime duration, which is longer at the lower PSSDR, and it's shorter at higher PSSDR. As yet SD $\sigma_{\text{scale}}^{\langle\max\rangle} = 0.2$ is sufficient to produce S6080I, watched on real practice, then the deal is just to set up the range of pixel distortion SD $\sigma_{\text{PD}}^{\langle\max\rangle}$. Continuing, the minimal-ultimate maximal pixel-distortion SD is $\sigma_{\text{PD}}^{\langle\max\rangle} = 0.002$ as by $\sigma_{\text{PD}}^{\langle\max\rangle} < 0.002$ values in the matrix $\boldsymbol{\Xi}$ or value $\xi\left(k\right)$ of NV with ZEUV are too insignificant, distorting pixels in S6080I sparsely. Then the lower boundary of PSSDR is $r_{\min} = 0.01$. Therefore, the upper boundary of PSSDR $r_{\max} = 1$ ensues from the maximal-ultimate maximal pixel-distortion SD is $\sigma_{\text{PD}}^{\langle\max\rangle} = 0.2$ as by $\sigma_{\text{PD}}^{\langle\max\rangle} > 0.2$ pixel distortions become unsuitable for training [1]. Consequently, the boundaries for PSSDR are evaluated as $r_{\min} = 0.01$ and $r_{\max} = 1$, and the segment of PSSDR is $\left[0.01;\, 1\right]$. Unlike HLNN being integer and having maximal quantity of points $N_{\text{HLN}}^{\langle\max\rangle} - N_{\text{HLN}}^{\langle\min\rangle} + 1$ from its range $\left[N_{\text{HLN}}^{\langle\min\rangle};\, N_{\text{HLN}}^{\langle\max\rangle}\right]$ to be run through, PSSDR has continuum of points in any $\left[r_{\min};\, r_{\max}\right]$ range by $r_{\max} > r_{\min}$. The step of sampling the range $\left[0.01;\, 1\right]$ isn't necessary to be equidistant: local minima may be probably discovered closer to the lower boundary $r_{\min} = 0.01$, and there the sampling should be denser. Re-sampling isn't excluded also. Well, these questions are coming to be highlighted in the next section.

9. RUNNING THROUGH THE CARTESIAN PRODUCT OF HLNN AND PSSDR RANGES

The Cartesian product of the abovementioned ranges of HLNN and PSSDR is horizontally-striped rectangle (HSR)

$$\left[r_{\min};\, r_{\max}\right] \times \left\{\left[N_{\text{HLN}}^{\langle\min\rangle};\, N_{\text{HLN}}^{\langle\max\rangle}\right] \cap \mathbb{N}\right\} = \left[0.01;\, 1\right] \times \left\{\left[100;\, 350\right] \cap \mathbb{N}\right\}. \tag{15}$$

Order of variables is surely out of importance. A two-component point from HSR (15)

$$\left[r \quad N_{\text{HLN}}\right] \in \left[r_{\min};\, r_{\max}\right] \times \left\{\left[N_{\text{HLN}}^{\langle\min\rangle};\, N_{\text{HLN}}^{\langle\max\rangle}\right] \cap \mathbb{N}\right\} = \left[0.01;\, 1\right] \times \left\{\left[100;\, 350\right] \cap \mathbb{N}\right\} \tag{16}$$

is the couple of 2LP parameters to be optimized. Mathematically, the optimization problem lies in minimizing and gripping the starred two-component point from (16) as the function minimum argument:

$$\left[r^* \quad N_{\text{HLN}}^* \right] \in \arg \left\{ \min_{\left[r \quad N_{\text{HLN}} \right] \in \left[r_{\min} ;\, r_{\max} \right] \times \left\{ \left[N_{\text{HLN}}^{\langle \min \rangle} ;\, N_{\text{HLN}}^{\langle \max \rangle} \right] \cap \mathbb{N} \right\}} p_{\text{error}} \left(r,\, N_{\text{HLN}} \right) \right\} . \tag{17}$$

In explicit form, the problem (17)

$$\left[r^* \quad N_{\text{HLN}}^* \right] \in \arg \left\{ \min_{\left[r \quad N_{\text{HLN}} \right] \in \left[0.01;\, 1 \right] \times \left\{ \left[100;\, 350 \right] \cap \mathbb{N} \right\}} p_{\text{error}} \left(r,\, N_{\text{HLN}} \right) \right\} \tag{18}$$

needs the function $p_{\text{error}} \left(r,\, N_{\text{HLN}} \right)$, evaluated on HSR (15) as a plane lattice, which is sampled from (15).

Primarily the range $\left[0.01;\, 1 \right]$ of PSSDR by $N_{\text{HLN}} = \overline{100,\, 350}$ should be run through for determining the shape of the averaged CEP $p_{\text{error}} \left(r,\, N_{\text{HLN}} \right)$ surface. HLNN is sampled roughly: the step is 10 neurons. Let the step within the range $\left[0.01;\, 1 \right]$ be 0.01 up to the point $r = 0.1$, and 0.1 up to the point $r = 1$, what gives 19 points within the segment $\left[0.01;\, 1 \right]$ to compute the averaged CEP in 494 points altogether (this is for one 2LP). Once local minima of the surface $p_{\text{error}} \left(r,\, N_{\text{HLN}} \right)$ are roughly determined, the range $\left[0.01;\, 1 \right]$ is narrowed to a subsegment of $\left[0.01;\, 1 \right]$ (the subsegment may be re-sampled with a smaller step). The sequence of HLNN within the segment $\left[100;\, 350 \right]$ can be narrowed as well.

Running through HSR (15) implies testing the PDS6080I-trained 2LP with S6080I. In the training process by PDS6080I, the input of 2LP is fed with the training set

$$\left\{ \tilde{\mathbf{P}}_i^{\langle \text{PDS6080I} \rangle} \right\}_{i=1}^{C+F} = \left\{ \left\{ \mathbf{A} \right\}_{l=1}^{C},\, \left\{ \mathbf{A}_{\text{PDS6080I}}^{\langle k \rangle} \right\}_{k=1}^{F} \right\} \tag{19}$$

of C replicas of all 26 classes' non-distorted representatives and F matrices of PDS6080I by the set of identifiers

$$\left\{ \mathbf{T}_i \right\}_{i=1}^{C+F} = \left\{ \mathbf{I} \right\}_{i=1}^{C+F} \tag{20}$$

with identity 26×26 matrix $\mathbf{I}$. The matrix $\mathbf{A}$ is formed by concatenating horizontally 4800-length-column-reshaped matrices $\left\{ \mathbf{A}_q \left(k \right) \right\}_{q=1}^{26}$. The set (19), being formed by (1) – (13) at some PSSDR (14), is passed through 2LP with identifiers (20) for Q_{pass} times.

For obtaining preliminarily the real tolerable CEP on average, it is sufficient to take parameters $C = 2$, $F = 8$, $Q_{\text{pass}} = 30$. The training set

$$\left\{ \tilde{\mathbf{P}}_i^{\langle \text{PDS6080I} \rangle} \right\}_{i=1}^{10} = \left\{ \left\{ \mathbf{A} \right\}_{l=1}^{2},\, \left\{ \mathbf{A}_{\text{PDS6080I}}^{\langle k \rangle} \right\}_{k=1}^{8} \right\} \tag{21}$$

feeds 2LP for 30 times with identifiers $\left\{ \mathbf{T}_i \right\}_{i=1}^{10}$, producing the PDS6080I-trained 2LP of performance that could be improved, though.

The PDS6080I-trained 2LP is tested with S6080I at the range of scale SD σ_{scale} from the minimal one up to $\sigma_{\text{scale}}^{\langle\max\rangle}$, that is $\sigma_{\text{scale}} \in [0;\, 0.2]$. The range $[0;\, 0.2]$ is going to be run through with the step 0.02, which lets evaluate CEP in those 11 different points of the scale SD.

Denote CEP of 2LP on S6080I by scale SD σ_{scale} as $p_{\text{error}}^{\langle\text{S6080I}\rangle}(r,\, N_{\text{HLN}};\, \sigma_{\text{scale}})$. Analytically, the averaged CEP

$$p_{\text{error}}(r,\, N_{\text{HLN}}) = \frac{1}{\sigma_{\text{scale}}^{\langle\max\rangle}} \int\limits_{\sigma_{\text{scale}} \in \left[0;\, \sigma_{\text{scale}}^{\langle\max\rangle}\right]} p_{\text{error}}^{\langle\text{S6080I}\rangle}(r,\, N_{\text{HLN}};\, \sigma_{\text{scale}})\, d\sigma_{\text{scale}} =$$

$$= 5\int_0^{0.2} p_{\text{error}}^{\langle\text{S6080I}\rangle}(r,\, N_{\text{HLN}};\, \sigma_{\text{scale}})\, d\sigma_{\text{scale}}\,. \tag{22}$$

And due to that the range of scale SD σ_{scale} is sampled with the step 0.02, statement (22) is computed as

$$p_{\text{error}}(r,\, N_{\text{HLN}}) = \frac{1}{11}\sum_{t=0}^{10} p_{\text{error}}^{\langle\text{S6080I}\rangle}(r,\, N_{\text{HLN}};\, 0.02t)\,. \tag{23}$$

Before beginning the run, 2LP is trained with the training set (21). Further, it is tested through the 494-pointed lattice

$$\left\{\{0.01+0.01n\}_{n=0}^{9},\, \{0.1+0.1m\}_{m=1}^{9}\right\} \times \left\{\{100+10h\}_{h=0}^{25}\right\} \subset$$

$$\subset [0.01;\, 1] \times \left\{[100;\, 350] \cap \mathbb{N}\right\} \tag{24}$$

of HSR (15), whereupon a rough mesh of the surface (23) can be drafted (fig. 1).

The local minimum at

$$[r \quad N_{\text{HLN}}] = [0.3 \quad 160] \tag{25}$$

is rather occasional, where $p_{\text{error}}(0.3,\, 160) < 0.7846$, indeed. By $r \in [0.1;\, 1]$ this is the single point, at which the averaged CEP is less than 0.8, and there are eight 2LP among those 18 ones with such a low CEP at (25). Altogether there are 12 points satisfying the condition $p_{\text{error}}(r,\, N_{\text{HLN}}) < 0.8$, and, except (25), the rest 11 ones are:

$$[r \quad N_{\text{HLN}}] = [0.01 \quad 110], \tag{26}$$

$$[r \quad N_{\text{HLN}}] = [0.02 \quad 240], \tag{27}$$

$$[r \quad N_{\text{HLN}}] = [0.03 \quad 140], \tag{28}$$

$$[r \quad N_{\text{HLN}}] = [0.03 \quad 150], \tag{29}$$

$$[r \quad N_{\text{HLN}}] = [0.03 \quad 230], \tag{30}$$

$$[r \quad N_{\text{HLN}}] = [0.03 \quad 320], \tag{31}$$

$$[r \quad N_{\text{HLN}}] = [0.03 \quad 350], \tag{32}$$

$$[r \quad N_{\text{HLN}}] = [0.04 \quad 130], \tag{33}$$

$$[r \quad N_{\text{HLN}}] = [0.05 \quad 150], \tag{34}$$

$$[r \quad N_{\text{HLN}}] = [0.07 \quad 140], \tag{35}$$

$$\begin{bmatrix} r & N_{\text{HLN}} \end{bmatrix} = \begin{bmatrix} 0.07 & 220 \end{bmatrix}. \tag{36}$$

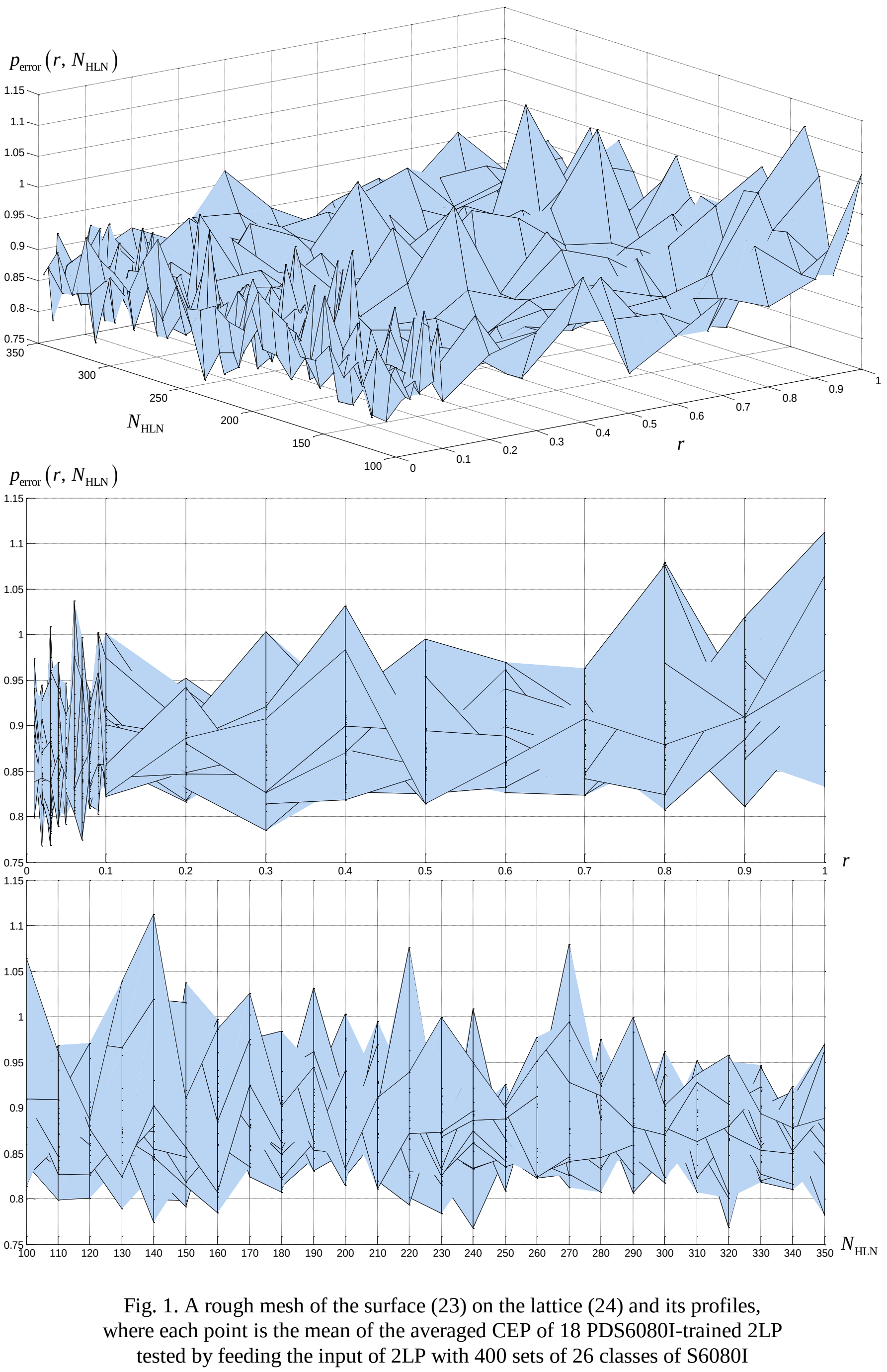

Fig. 1. A rough mesh of the surface (23) on the lattice (24) and its profiles, where each point is the mean of the averaged CEP of 18 PDS6080I-trained 2LP tested by feeding the input of 2LP with 400 sets of 26 classes of S6080I

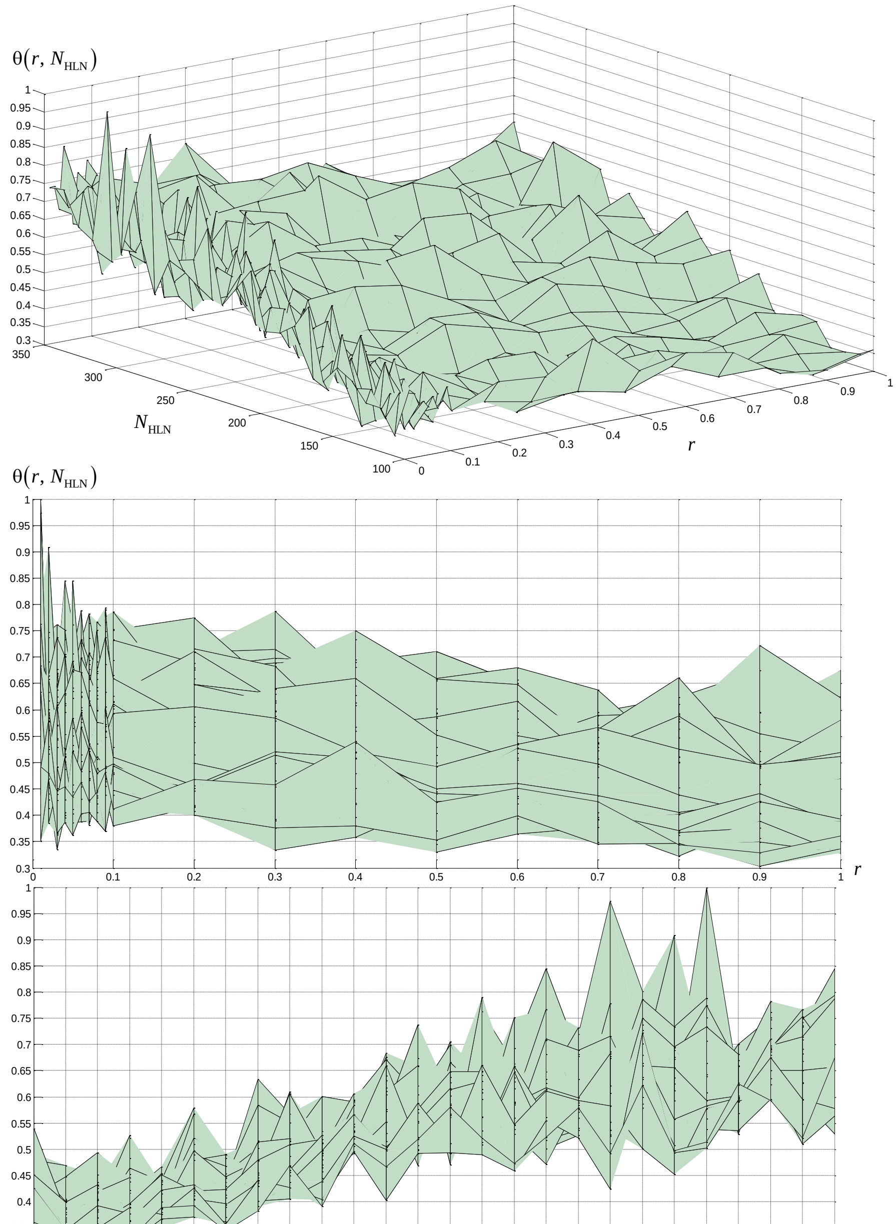

Fig. 2. An evaluation of the unit-normed traintime surface on the lattice (24) and its profiles, where each point is the mean of the unit-normed traintimes of 18 PDS6080I-trained 2LP tested for fig. 1

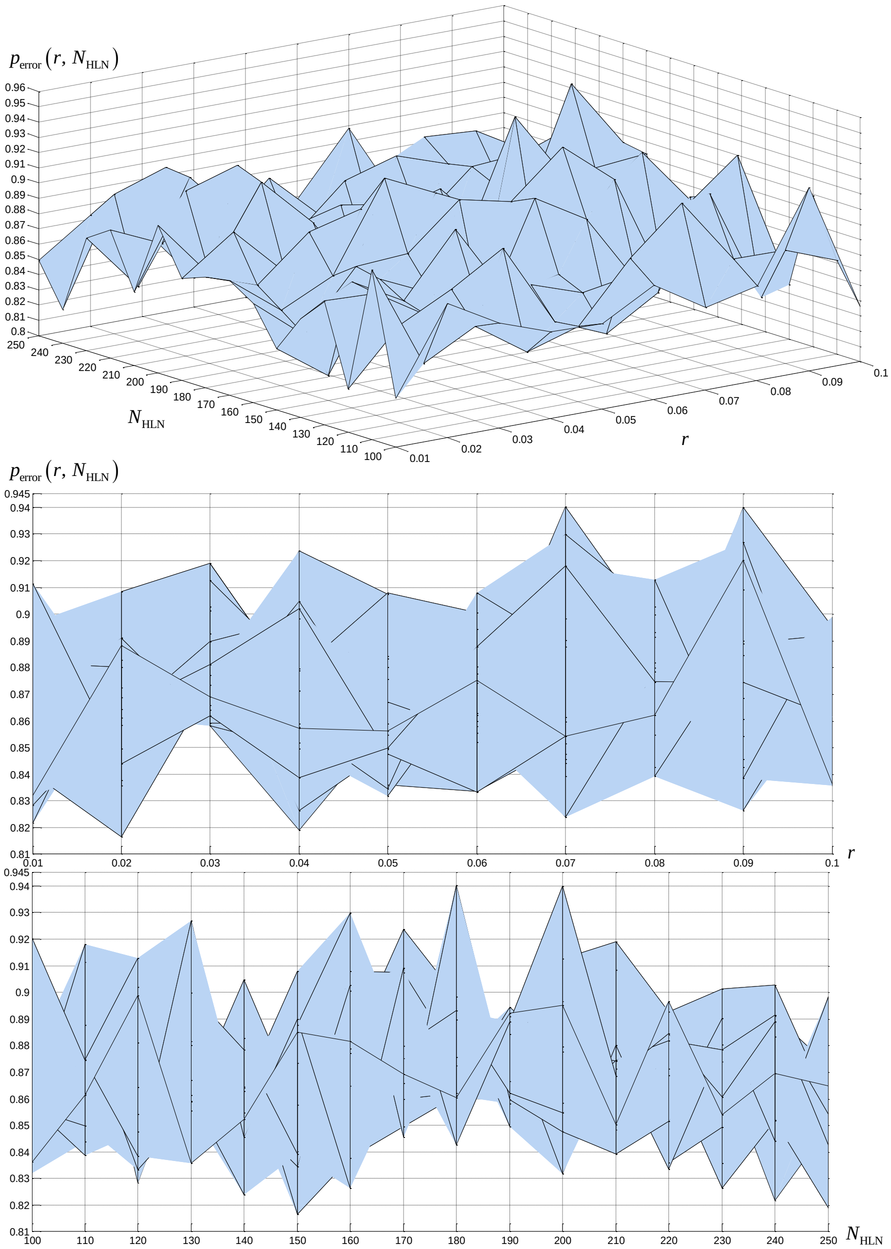

Fig. 3. A locally refined mesh of the surface (23) on the lattice (38) and its profiles, where each point is the mean of the averaged CEP of 70 PDS6080I-trained 2LP tested by feeding the input of 2LP with 400 sets of 26 classes of S6080I (in each point, the testing results of the previously 18 PDS6080I-trained 2LP are included)

All the points (26) – (36) have $r < 0.1$, and only two points (31) and (32) have $N_{\text{HLN}} > 250$. These two points are paid attention to, because $p_{\text{error}}(0.03, 320) < 0.7685$ and $p_{\text{error}}(0.03, 350) < 0.7821$, and there are 21 PDS6080I-trained 2LP among those 36 ones (11 and 10 ones among each of 18, respectively) whose averaged CEP is less than 0.8. Nevertheless, $p_{\text{error}}(0.03, 140) < 0.7995$ and $p_{\text{error}}(0.03, 150) < 0.7982$, where HLNN is far less, and there are 20 PDS6080I-trained 2LP among those 36 ones (11 and 9 ones among each of 18, respectively) whose averaged CEP is less than 0.8. And greater HLNN requires longer traintime – see how the unit-normed traintime $\theta(r, N_{\text{HLN}})$ in fig. 2 varies on the lattice (24). So, the points (31) and (32) are excluded, and HSR (15) is narrowed to a smaller HSR

$$[0.01;\ 0.1] \times \{[100;\ 250] \cap \mathbb{N}\} \subset [0.01;\ 1] \times \{[100;\ 350] \cap \mathbb{N}\}. \tag{37}$$

Owing to the narrowing, PDS6080I-trained 2LP is tested through the 160-pointed lattice

$$\left\{\{0.01 + 0.01n\}_{n=0}^{9}\right\} \times \left\{\{100 + 10h\}_{h=0}^{15}\right\} \subset [0.01;\ 0.1] \times \{[100;\ 250] \cap \mathbb{N}\} \subset$$
$$\subset [0.01;\ 1] \times \{[100;\ 350] \cap \mathbb{N}\} \tag{38}$$

of HSR (37), but now each point is evaluated over 70 PDS6080I-trained 2LP (fig. 3).

The locally refined mesh in fig. 3 has two local minima at

$$[r \quad N_{\text{HLN}}] = [0.02 \quad 150] \tag{39}$$

and

$$[r \quad N_{\text{HLN}}] = [0.04 \quad 250] \tag{40}$$

with very low CEP: $p_{\text{error}}(0.02, 150) < 0.8165$ and $p_{\text{error}}(0.04, 250) < 0.819$, respectively. Here we get slightly greater CEP because of averaging over 70 PDS6080I-trained 2LP instead of 18 ones (fig. 4). Thus, the scattering of the averaged CEP decreases.

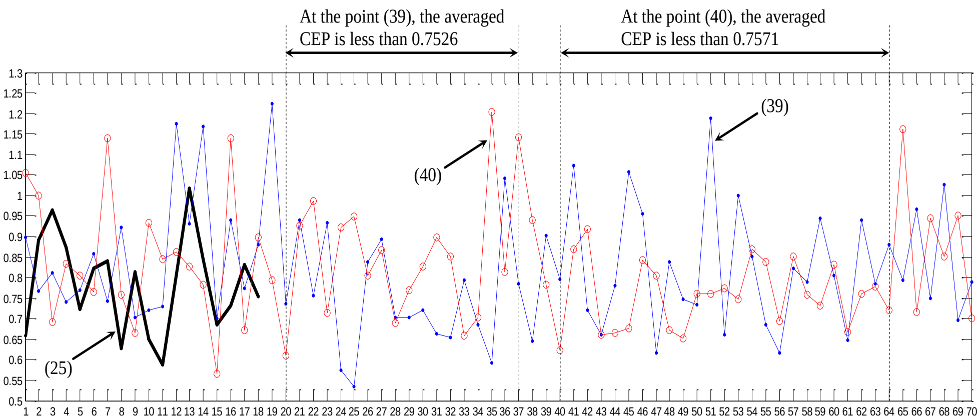

Fig. 4. The averaged CEP scatter polylines disclosed for points (25), (39), (40), where abscissa axis is a number of PDS6080I-trained 2LP; the averaged CEP at (39) reaches 0.7526 if it is calculated over the 18-pointed window starting off the 20-th point, and the averaged CEP at (40) reaches 0.7571 if it is calculated over the 25-pointed window starting off the 40-th point

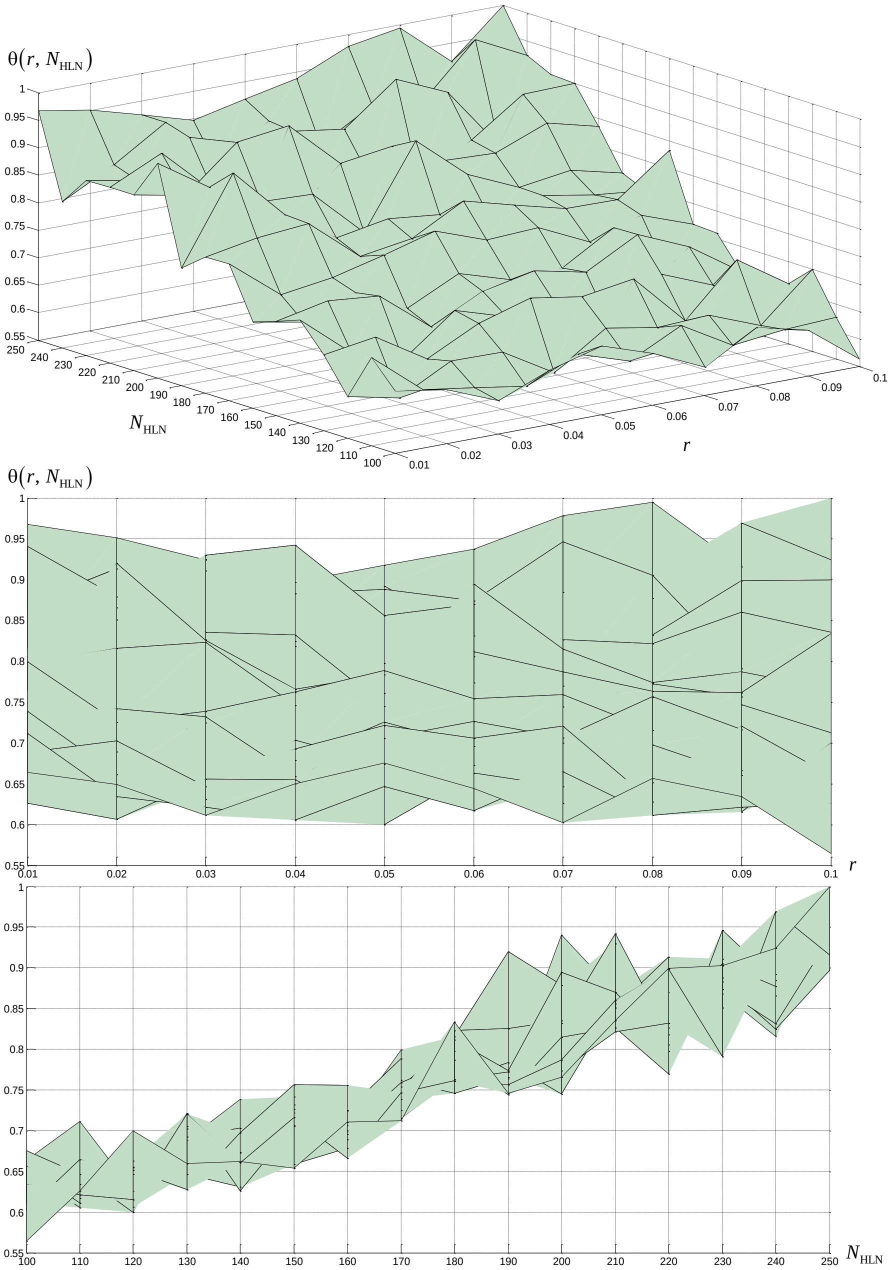

Fig. 5. A re-evaluation of the unit-normed traintime surface on the lattice (38) and its profiles, where each point is the mean of the unit-normed traintimes of 70 PDS6080I-trained 2LP tested for fig. 3

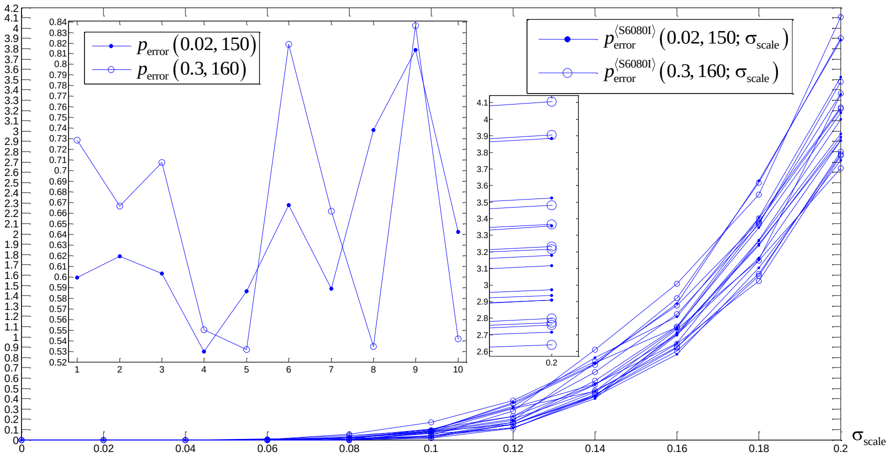

Fig. 6. The averaged performance and SD $\{0.02t\}_{t=0}^{10}$ -disclosed performance of 10 PDS6080I-trained 2LP by the points (25) and (39); each 2LP is tested by feeding the input of 2LP with 800 sets of 26 classes of S6080I

In selecting between the points (39) and (40), the unit-normed traintime helps again. Fig. 5 prompts that selection of the point (39) is appropriate owing to its traintime is almost 40 % shorter. Another finding in favor of the point (39) is simplicity of the classifier, where the lesser HLNN corresponds to the simpler 2LP.

Eventually, the points (25) and (39) remain pretending to be solutions of the problem (18). Due to 10 neurons difference, advantage of the point (39) is minor. To ascertain the global minimum of the surface $p_{\text{error}}(r, N_{\text{HLN}})$ as the single solution, 2LP shall be trained and tested at the points (25) and (39) more punctiliously.

10. SOLUTION OF THE PROBLEM (18) AND ITS VERIFICATION

The preliminarily reached tolerable CEP (about 0.82) can be decreased if pass the training set (21) through 2LP longer. So take $Q_{\text{pass}} = 50$ and increase the testing sets' number of 26 classes of S6080I feeding the input of 2LP up to 800. Then it turns out, that 2LP in the point (39) is trained better (fig. 6).

By the data to fig. 6, the mean of the averaged CEP of 10 PDS6080I-trained 2LP in the point (39) is slightly lesser:

$$p_{\text{error}}(0.02, 150) < 0.6389 < 0.6577 < p_{\text{error}}(0.3, 160). \tag{41}$$

Anyway, according to (41) and fig. 3 and 5, the point (39) is treated as the solution of the problem (18). Hence,

$$\left[r^{*} \quad N_{\text{HLN}}^{*} \right] = \left[0.02 \quad 150 \right] \tag{42}$$

and, for verification, let us see the performance within an HSR neighborhood, containing the point (42). We don't need to break the lattice (38) denser, because stochasticity is very high, and the solution is going to be approximate through both axes.

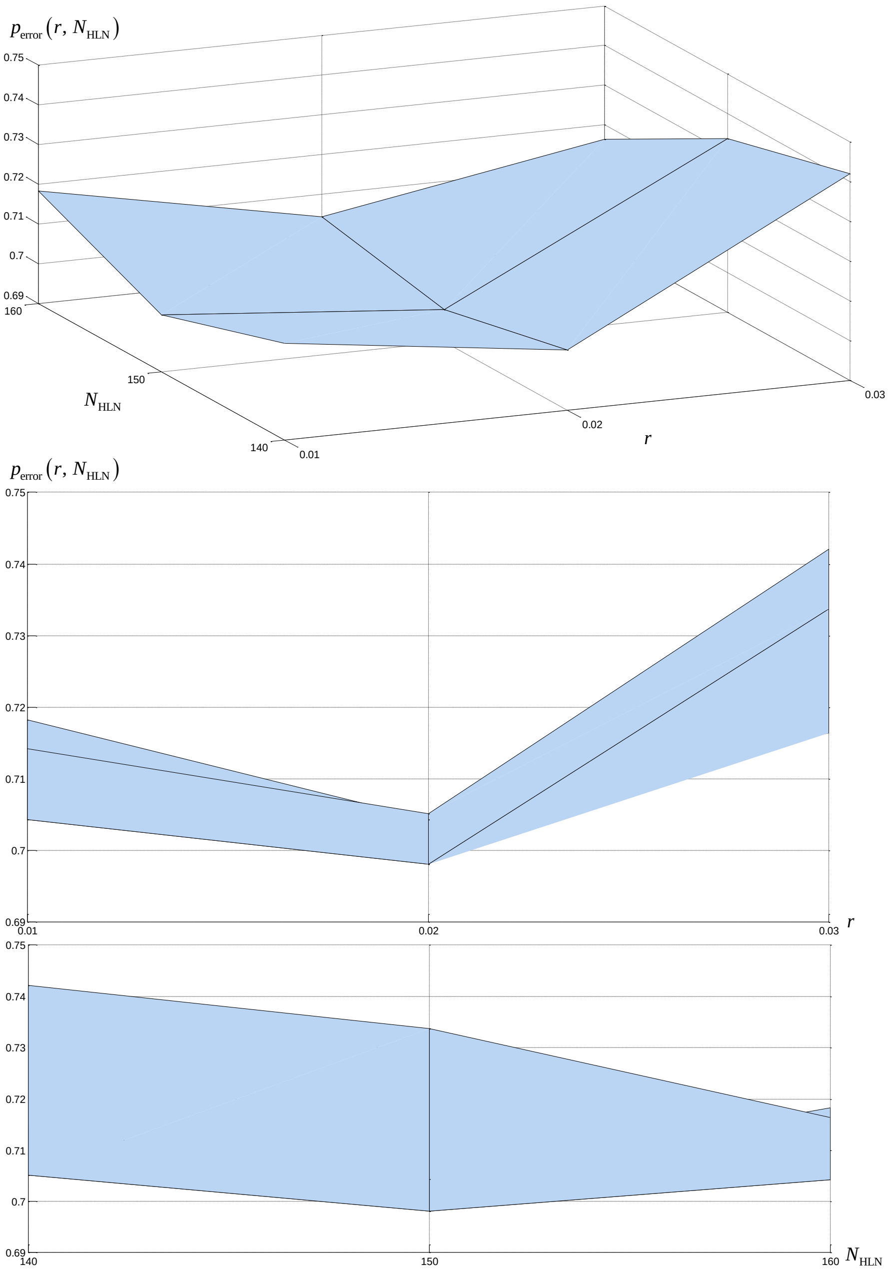

Fig. 7. A re-evaluation of the surface (23) mesh on the lattice (43) and its profiles, where each point is the mean of the averaged CEP of 200 PDS6080I-trained 2LP by $Q_{\text{pass}} = 50$ tested with 800 sets of 26 classes of S6080I

Therefore, 2LP ought to be trained by nine points of the 3×3 lattice

$$\{0.01,\ 0.02,\ 0.03\}\times\{140,\ 150,\ 160\}\subset[0.01;\ 0.1]\times\{[100;\ 250]\cap\mathbb{N}\}\subset$$
$$\subset[0.01;\ 1]\times\{[100;\ 350]\cap\mathbb{N}\}\ . \tag{43}$$

But the point (39) is already re-evaluated by $Q_{\text{pass}} = 50$ and 800 testing sets of 26 classes of S6080I (for fig. 6), and there remain eight points.

Fig. 7 confirms optimality of the point (42). Performance of the best PDS6080I-trained 2LP at this point is

$$p_{\text{error}}(0.02,\ 150) < 0.4507\ . \tag{44}$$

However, there is even better performance which is $p_{\text{error}}(0.01,\ 160) < 0.4022$, though. But this incongruity is just an outcome of stochasticity and small percentages. And, statistically, mathematical expectation of CEP at the point (42) is the closest to minimum.

As the solution (42) of the problem (18) has been verified, now the question is about PDS6080I. Denote CEP of 2LP on PDS6080I by scale SD σ_{scale} and $r = 1$ (this PSSDR relates to the testing sets) as $p_{\text{error}}^{\langle \text{PDS6080I} \rangle}(r,\ N_{\text{HLN}};\ \sigma_{\text{scale}})$. Analytically, the averaged CEP

$$p_{\text{error}}(r,\ N_{\text{HLN}}) = 5\int_{0}^{0.2} p_{\text{error}}^{\langle \text{PDS6080I} \rangle}(r,\ N_{\text{HLN}};\ \sigma_{\text{scale}})\,d\sigma_{\text{scale}} \tag{45}$$

is approximated to

$$p_{\text{error}}(r,\ N_{\text{HLN}}) = \frac{1}{11}\sum_{t=0}^{10} p_{\text{error}}^{\langle \text{PDS6080I} \rangle}(r,\ N_{\text{HLN}};\ 0.02t) \tag{46}$$

like the averaged CEP in (22) and (23). The optimized PDS6080I-trained 2LP maintains its capability to classify PDS6080I accurately:

$$p_{\text{error}}(0.02,\ 150) < 0.5582 \ \text{ and } \ p_{\text{error}}(0.01,\ 160) < 0.5097\ . \tag{47}$$

The best PDS6080I-trained at $\left\{\sigma_{\text{scale}}^{\langle \max \rangle} = 0.2,\ \sigma_{\text{PD}}^{\langle \max \rangle} = 0.004\right\}$ 2LP by the training set (21) has been tested at the highest scale SD $\sigma_{\text{scale}} = 0.2$ for S6080I and at $r = 1$ for PDS6080I, respectively. Fig. 8 shows of how much the monochrome image is distorted when it's been scaled at SD $\sigma_{\text{scale}} = 0.2$, and nonetheless S6080I or PDS6080I is classified excellently. The mere one letter from so-scaled 37 objects is classified wrong. One letter PDS6080I is classified wrong from 35 objects.

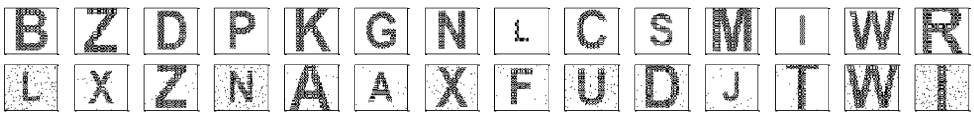

Fig. 8. S6080I at scale SD $\sigma_{\text{scale}} = 0.2$ and PDS6080I (beneath) at scale SD $\sigma_{\text{scale}} = 0.2$ by $r = 1$, where the mere one letter from so-scaled 37 objects is classified wrong by the PDS6080I-trained 2LP with optimal HLNN and PSSDR

At the average, the one letter from the 220 scaled objects is classified wrong. If the scaling amplitude is slight ($\sigma_{\text{scale}} < 0.05$), possibility of wrong classification is negligibly small. Assuredly, non-scaled objects are classified correctly ever.

11. CONCLUSION AND SUGGESTIONS FOR FURTHER 2LP PERFORMANCE OPTIMIZATION

This article is a show of how neurons number in the perceptron and an element of the topological configuration for training it can be optimized and verified for the definite classification problem over a fixed pattern of objects. The results of this article lie in the shown way of 2LP performance optimization over its two parameters. The optimization process goes on while the classifier is identified. This process concerned CEP solely, without referring to traintime. In furthering, the other parameters can be optimized to decrease the averaged CEP more. For 2LP, they are the number C for the training set (19); the number F for (19), indicating at smoothness in the training process with (19) by (20) – the greater this number, the smoother training process is; the number Q_{pass} (if passing the training set through 2LP is strictly limited). Being subsidiary, the traintime duration could be optimized if it is supposed to be limited when the classifier is constrained everywhen to become fine-adjusted to a new general totality of the previous type objects. Optimums, however, clearly may vary when type of objects changes from the monochrome image to the colored or elsewise. The same happens when number of the object features changes (the image is re-formatted).

It is obvious the problem (18) can be solved only numerically, analyzing the mesh (23). And another open question is how to speed up the process of evaluating the function (23) and similar ones. Parallelization techniques [5] are believable to serve it.

REFERENCES

1. *Romanuke V. V.* Pixel-to-scale standard deviations ratio optimization for two-layer perceptron training on pixel-distorted scaled 60-by-80-images in scaled objects classification problem / V. V. Romanuke // Вісник Кременчуцького національного університету імені М. Остроградського. – 2014. – Вип. 2 (85). – С. 96–105.
2. *You J.* Classification and segmentation of rotated and scaled textured images using texture “tuned” masks / J. You, H. A. Cohen // Pattern Recognition. – 1993. – Vol. 26, Iss. 2. – P. 245–258.
3. *Barat C.* String representations and distances in deep Convolutional Neural Networks for image classification / C. Barat, C. Ducottet // Pattern Recognition. – 2016. – Vol. 54. – P. 104–115.
4. *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 / V. V. Romanuke // Наукові вісті НТУУ “КПІ”. – 2014. – № 5. – С. 55–62.
5. *Lee J.* Adaptive execution techniques of parallel programs for multiprocessors / J. Lee, J.-H. Park, H. Kim, C. Jung, D. Lim, S. Han // Journal of Parallel and Distributed Computing. – 2010. – Vol. 70, Iss. 5. – P. 467–480.

Стаття: надійшла до редколегії 03.05.2016
доопрацьована 25.01.2017
прийнята до друку 15.03.2017

ОПТИМАЛЬНЕ ЧИСЛО НЕЙРОНІВ ПРИХОВАНОГО ШАРУ ДВОШАРОВОГО ПЕРСЕПТРОНА ТА СПІВВІДНОШЕННЯ СКВ ПІКСЕЛЬНИХ СПОТВОРЕНЬ І МАСШТАБУВАННЯ ДЛЯ ЙОГО НАВЧАННЯ НА МАСШТАБОВАНИХ ЗОБРАЖЕННЯХ ФОРМАТУ 60-НА-80 З ПІКСЕЛЬНИМИ СПОТВОРЕННЯМИ У ЗАДАЧІ КЛАСИФІКАЦІЇ МАСШТАБОВАНИХ ОБ'ЄКТІВ

В. Романюк

Хмельницький національний університет,
вул. Інститутська, 11, м. Хмельницький, 29016

Розглянуто задачу класифікації. Об'єкти для класифікації спотворені з ефектом лінійного масштабування. Класифікатором є двошаровий персептрон. Модель об'єкта – монохромне 60-на-80-зображення збільшеної великої літери англійського алфавіту. Відтак генеральна сукупність сформована з монохромних 60-на-80-зображень алфавітних літер, становлячи 26 класів. Наша мета – довести, як число нейронів прихованого шару у двошаровому персептроні й елемент топологічної конфігурації для його навчання можуть бути оптимізовані для задачі класифікації масштабованих об'єктів. Цим елементом топологічної конфігурації є співвідношення СКВ піксельних спотворень і масштабування. Це співвідношення дає підстави додавати у навчальну множину об'єкти зі спотвореними ознаками. Це спотворення ознак формується через додавання нормального шуму з нульовим середнім і дисперсією, яка регулюється цим співвідношенням. Двошаровий персептрон моделюється, навчається та тестується у MATLAB. Фактично, це оптимізаційна задача з двома змінними, де продуктивність персептрона оптимізується у сенсі зменшення відсотка помилок класифікації. Він є функцією, яку оцінюють на декартовому добутку діапазонів числа нейронів прихованого шару та співвідношення СКВ. Цей добуток прямокутник з горизонтальними смугами, який у подальшому дискретизується до гратки. Результатом оптимізації є 150 нейронів у прихованому шарі та співвідношення, що дорівнює 0.02, які дають змогу отримувати класифікатор, де лише одна помилка трапляється на 37 об'єктів з максимальним ефектом масштабування.

Ключові слова: класифікація масштабованих об'єктів, двошаровий персептрон, монохромне зображення, навчальна множина, відсоток помилок класифікації, оптимальне число нейронів прихованого шару, оптимальне співвідношення середньоквадратичних відхилень піксельних спотворень і масштабування.

ISSN 2078–5097. Вісн. Львів. ун-ту. Сер. прикл. матем. та інф. 2017. Вип. 25. С. 130–130
Visnyk of the Lviv University. Series Appl. Math. and Informatics. Issue 25. P. 130–130