UDC 004.032.26::004.93+519.8

PIXEL-TO-TURN STANDARD DEVIATIONS RATIO OPTIMIZATION FOR TRAINING TWO-LAYER PERCEPTRON ON PIXEL-DISTORTED TURNED 60-BY-80-IMAGES IN TURNED OBJECTS CLASSIFICATION PROBLEM

Romanuke V. V., c. t. s., associate professor

Khmelnitskiy National University,
Institutskaya str., 11, Khmelnitskiy, 29016, Ukraine

romanukevadimv@mail.ru

The problem of turned objects classification is reviewed. This problem partially issues from that there cannot be ensured the symmetry just on the moment when the object is captured after it has been traced. On the other hand, sometimes it is harder to symmetrize the object than to centralize it. The classifier must identify the object to its definite class, without regard to the rotation or angulation span. Hierarchical multilayered neural networks, cognitrons, neocognitrons and convolutional neural networks perform perfectly over turned objects for their classification, but they take long periods for the performance and grab gigantic information-memory resources. Perceptrons being universal classifiers with having one or more hidden layers consume far less information-memory resources, and they are much faster. However, perceptrons perform perfectly just over the object whose features are distorted by a statistical law, which is near to normal distribution. The turn distortion is far off the normal distribution and, factually, turned objects classification is either expensive or ineffective. Thus, improvement of the classification process with the classifier training parameters optimization is considered. The classifier type is two-layer perceptron. The object of classification is monochrome 60-by-80-image of the enlarged English alphabet capital letter, which may be distorted with turns and, rarer, be pixel-distorted. Therefore, general totality is of 26 classes. The goal is to find a pixel-to-turn standard deviations ratio for the training process in order to ensure minimum of classification error percentage. The two-layer perceptron is modeled, trained and tested under 250 neurons in its hidden layer and «traingda» MATLAB training function. After having determined the endpoints of the range of pixel-to-turn standard deviations ratio, there are trained many series of two-layer perceptrons over this range, whereupon they are batch-tested. The minimum of classification error percentage is sought over polylines, approximating its dependence on the ratio. At that, the optimal ratio should fit as for classifying turned images, as well as for classifying pixel-distorted turned images. After having accumulated huge statistical data, there is found the optimal ratio, which allows to train two-layer perceptron classifiers optimally. A best-trained classifier produces classification error percentage not greater than 0.19 % over turn-distorted images, and not greater than 0.29 % over pixel-distorted turned images. For other general totalities, or other number of classes, or other object model the optimal ratio may be different, but goal and tasks within the third section of this article sustain the algorithm of how make the optimization in general.

Key words: turned objects classification, classification error percentage, two-layer perceptron, training set, MATLAB function «traingda», monochrome image, pixel-to-turn standard deviations ratio optimization.

ОПТИМИЗАЦИЯ СООТНОШЕНИЯ СРЕДНЕКВАДРАТИЧНЫХ ОТКЛОНЕНИЙ ПИКСЕЛЬНЫХ ИСКАЖЕНИЙ И ПОВОРОТОВ ДЛЯ ОБУЧЕНИЯ ДВУХСЛОЙНОГО ПЕРСЕПТРОНА НА ПОВЁРНУТЫХ ИЗОБРАЖЕНИЯХ ФОРМАТА 60-НА-80 С ПИКСЕЛЬНЫМИ ИСКАЖЕНИЯМИ В ЗАДАЧЕ КЛАССИФИКАЦИИ ОБЪЕКТОВ С ПОВОРОТАМИ

Романюк В. В., к. т. н., доцент

Хмельницкий национальный университет,
ул. Институтская, 11, г. Хмельницкий, 29016, Украина

romanukevadimv@mail.ru

Анализируется задача классификации объектов с поворотами. Эта задача отчасти исходит из того, что симметрия не может быть обеспечена именно в тот момент, когда объект попадает в поле захвата после того, как он был отслежен. С другой стороны, временами сложнее объект симметризировать, чем центрировать. Классификатор должен идентифицировать объект к его определённому классу, не принимая во внимание протяжённость вращения или установки под углом. Иерархические многослойные нейронные сети, когнитроны, неокогнитроны и нейронные сети свёрточного типа функционируют идеально для классификации объектов с поворотами, но им нужны длительные периоды для функционирования и они захватывают гигантские информационные ресурсы и память. Персептроны, будучи универсальными классификаторами с

одним или несколькими скрытыми слоями, обходятся гораздо меньшими информационными ресурсами и памятью, и они значительно быстрее. Однако, персептроны функционируют идеально только для классификации объектов, чьи признаки искажаются в соответствии с неким статистическим законом, который близок к нормальному распределению. Искажение поворотами далеко от нормального распределения, и классификация объектов с поворотами является фактически либо затратной, либо неэффективной. Соответственно и рассматривается улучшение процесса классификации с помощью оптимизации параметров обучения классификатора. Типом классификатора является двухслойный персептрон. Объектом классификации является монохромное изображение формата 60-на-80 увеличенной заглавной буквы английского алфавита, которое может быть искажено поворотами и, реже, быть с пиксельными искажениями. Поэтому генеральная совокупность состоит из 26 классов. Целью является нахождение определённого соотношения среднеквадратичных отклонений пиксельных искажений и поворотов в процессе обучения для того, чтобы обеспечить минимум процента ошибок классификации. Двухслойный персептрон моделируется, обучается и тестируется при 250 нейронах в его скрытом слое и MATLAB-функции для обучения «traingda». После определения концов диапазона соотношения среднеквадратичных отклонений пиксельных искажений и поворотов, много последовательностей двухслойных персептронов в этом диапазоне обучаются, после чего они тестируются в пакетном режиме. Минимум процента ошибок классификации ищется на ломаных, которые аппроксимируют его зависимость от данного соотношения. При этом оптимальное соотношение должно подходить как для классификации изображений с поворотами, так и для классификации повёрнутых изображений с пиксельными искажениями. После накопления достаточно большой совокупности статистических данных, находится оптимальное соотношение, которое позволяет оптимально обучать классификаторы на основе двухслойных персептронов. Один из наилучших обученных классификаторов функционирует с процентом ошибок классификации, который не превышает 0.19 % для изображений, искажённых поворотами, и не превышает 0.29 % для повёрнутых изображений с пиксельными искажениями. Для иных генеральных совокупностей, или иного количества классов, или иной модели объекта данное оптимальное соотношение может отличаться, однако цель и задачи в пределах третьего раздела настоящей статьи поддерживают алгоритм того, как выполнять оптимизацию в общем.

Ключевые слова: классификация объектов с поворотами, процент ошибок классификации, двухслойный персептрон, обучающее множество, MATLAB-функция «traingda», монохромное изображение, оптимизация соотношения среднеквадратичных отклонений пиксельных искажений и поворотов.

ОПТИМІЗАЦІЯ СПІВВІДНОШЕННЯ СЕРЕДНЬОКВАДРАТИЧНИХ ВІДХИЛЕНЬ ПІКСЕЛЬНИХ СПОТВОРЕНЬ І ПОВОРОТІВ ДЛЯ НАВЧАННЯ ДВОШАРОВОГО ПЕРСЕПТРОНУ НА ПОВЕРНУТИХ ЗОБРАЖЕННЯХ ФОРМАТУ 60-НА-80 З ПІКСЕЛЬНИМИ СПОТВОРЕННЯМИ В ЗАДАЧІ КЛАСИФІКАЦІЇ ОБ'ЄКТІВ З ПОВОРОТАМИ

Романюк В. В., к. т. н., доцент

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

romanukevadimv@mail.ru

Аналізується задача класифікації об'єктів з поворотами. Ця задача частково походить з того, що симетрія не може бути забезпечена саме в той момент, коли об'єкт потрапляє у поле захвата після того, як він був відстежений. З іншого боку, часом складніше об'єкт симетризувати, ніж центрувати. Класифікатор повинен ідентифікувати об'єкт до його певного класу, не приймаючи до уваги довжину обертання чи установки під кутом. Ієрархічні багатошарові нейронні мережі, когнітрони, неокогнітрони та конволютивні нейронні мережі функціонують ідеально для класифікації об'єктів з поворотами, але вони потребують тривалих періодів для функціонування та захоплюють гігантскі інформаційні ресурси і пам'ять. Персептрони, будучи універсальними класифікаторами з одним або декількома прихованими шарами, обходяться набагато меншими інформаційними ресурсами і пам'яттю, и вони значно швидші. Однак персептрони функціонують ідеально лише для класифікації об'єктів, чиї ознаки спотворюються відповідно до якогось статистичного закону, який близький до нормального розподілу. Спотворення поворотами є далеким від нормального розподілу, і класифікація об'єктів з поворотами є фактично або витратною, або неефективною. Відповідно і розглядається покращення процесу класифікації за допомогою оптимізації параметрів навчання класифікатора. Типом класифікатора є двошаровий персептрон. Об'єктом класифікації є монохромне зображення формату 60-на-80 збільшеної великої літери англійського алфавіту, котре може бути спотвореним поворотами і, рідше, бути з піксельними спотвореннями. Відтак генеральна сукупність складається з 26 класів. Метою є знаходження певного співвідношення середньоквадратичних відхилень піксельних спотворень і поворотів у процесі навчання для того, щоб забезпечити мінімум відсотка помилок класифікації. Двошаровий персептрон моделюється, навчається та тестується за 250 нейронів у його прихованому шарі та MATLAB-функції для навчання «traingda». Після визначення кінців діапазону співвідношення середньоквадратичних відхилень піксельних спотворень і

поворотів багато послідовностей двошарових персептронів у цьому діапазоні навчаються, після чого вони тестуються в пакетному режимі. Мінімум відсотка помилок класифікації шукається на ламаних, що апроксимують його залежність від даного співвідношення. При цьому оптимальне співвідношення має підходити як для класифікації зображень з поворотами, так і для класифікації повернутих зображень з піксельними спотвореннями. Після накопичення чималої сукупності статистичних даних знаходиться оптимальне співвідношення, котре дозволяє оптимально навчати класифікатори на основі двошарових персептронів. Один з найкращих навчених класифікаторів функціонує з відсотком помилок класифікації, що не перевищує 0.19 % для зображень, спотворених поворотами, і не перевищує 0.29 % для повернутих зображень з піксельними спотвореннями. Для інших генеральних сукупностей, або іншої кількості класів, або іншої моделі об'єкта дане оптимальне співвідношення може різнитися, однак мета і задачі в межах третього розділу цієї статті підтримують алгоритм того, як виконувати оптимізацію загалом.
Ключові слова: класифікація об'єктів з поворотами, відсоток помилок класифікації, двошаровий персептрон, навчальна множина, MATLAB-функція «traingda», монохромне зображення, оптимізація співвідношення середньоквадратичних відхилень піксельних спотворень і поворотів.

PROBLEM OF TURNED OBJECTS CLASSIFICATION

In classifying objects like when recognizing plane images or their contours the object rotation effect springs up often and often. It is caused with that there cannot be ensured the symmetry just on the moment when the object is captured after it has been traced. And sometimes it is harder to symmetrize the object than to centralize it. Without regard to the rotation or angulation span, the classifier must identify it to its definite class. The angle of rotation in the turned image has no matter.

A hard problem of turned objects classification is the classifier operation speed by reducing the huge information-memory resources (IMR) which are consumed by the classification system, including the classifier [1, 2]. This is because those hierarchical multilayered neural networks, cognitrons, neocognitrons and convolutional neural networks, performing perfectly over turned objects for their classification, take long periods for the performance and grab gigantic IMR [3, 4]. Perceptrons (with one or more hidden layers) consume far less IMR, and they are much faster. Two-layer perceptron (2LP), that is perceptron with the single hidden layer (whose neurons number can be varied and adjusted), is the fastest for most of classification problems, where the object features are distorted by a statistical law (especially if it is near-normal distribution). And if 2LP could be trained on turned objects, it would become a smart classifier over them, surpassing neocognitrons and others above.

IMPROVING THE CLASSIFICATION PROCESS WITH THE CLASSIFIER PARAMETERS OPTIMIZATION

For classification process improvement the classifier parameters must be optimized before running its training process. Actually, the classifier parameters optimization is only possible before it is trained (being under supervised learning). Once 2LP has been trained, the classification process attains its characteristics, whose stable estimations can be valued via statistics of batch testings of the trained 2LP. The training process for 2LP has a great many of its parameters [5, 6]. Excluding parameters of 2LP itself, the training process is defined with the training set, feeding the input of 2LP. Parameters of the training set in the training process are specific for a classification problem, being stirred up with the object type and types of objects distortion (shift, turn, severe rotation, skewness, scale, nonlinear distortion, etc.). Particularly, in the training process for classifying turned images the training set is formed as addition of matrix of pixel-distorted images and matrix of turned images. Intensity (deepness) of pixel distortion (PD) is measured via PD standard deviation (SD) σ_{PD} and intensity of turn distortion (TD) is measured via SD σ_{TD}, determining the range of left and right angulation due to TD. And an important question is that what ratio $r = \dfrac{\sigma_{\text{PD}}}{\sigma_{\text{TD}}}$ shall be here. It is certain that by the substantiated boundaries $r_{\min}$ and $r_{\max}$ for pixel-to-turn standard deviations ratio (PTSDR) r, enclosing the range $\left[r_{\min}; r_{\max}\right]$ of PTSDR permissible values, the optimal PTSDR is

$$r^* \in \arg \min_{r \in [r_{\min};\, r_{\max}]} p_{\mathrm{EC}}(r) \qquad (1)$$

by the function $p_{\mathrm{EC}}(r)$, whose value at PTSDR r is an averaged classification error percentage (CEP) over turned images (objects, in general). The training set, formed with PTSDR (1), will feed the input of 2LP until it becomes a better classifier due to minimizing in (1). Of course, the better classifier the more improved classification process is.

GOAL AND TASKS

Before solving the minimization problem (1) the function $p_{\mathrm{EC}}(r)$ must be evaluated. The minimization problem (1) is the goal, after which 2LP can be trained and perform smarter, surpassing some other types of classifiers. The function $p_{\mathrm{EC}}(r)$ evaluation requires the object type (format), number of classes, non-distorted representatives (NDR) of classes in the being defined general totality (GT). These tasks will be fulfilled in the next section. Then there goes a model of 2LP and its configuration. After that boundaries $r_{\min}$ and $r_{\max}$ for PTSDR, enclosing the range $[r_{\min}; r_{\max}]$ of PTSDR, are substantiated, based on the object type (format). Therein the range $[r_{\min}; r_{\max}]$ is sampled with steps for running through this range as fast as possible, but the results would be adequate. Having run through the range $[r_{\min}; r_{\max}]$ and plotted an averaged CEP $p_{\mathrm{EC}}(r)$ over this range solves the problem (1) numerically (graphically) on the statistics of the trained 2LP, achieving the goal eventually. The tasking will be completed by verifying whether CEP $p_{\mathrm{EC}}(r^*)$ is minimal over the modeled turned objects and inferring from the investigation results.

NDR OF CLASSES IN GT

Let's put the object to be classified as the monochrome image (MI) of 60×80 format. It's a convenient object model because MI distortions can be watched and rendered readily. The being suggested MI format isn't small and large, it's about medium, having 4800 features as the object for classification. This will ensure the classification results good theoretical, with power to be propagated on other turned objects, and investigation procedures and the classification results acquisition will be running without delays.

May the file format be bitmap. Within MATLAB, having powerful Neural Network Toolbox [7, 8] for programming and simulating neural networks, MI bitmap file is coded with ones (white color) and zeros (black color). Therefore the finite GT is of 60×80 matrices of ones and zeros (MOZ), containing altogether 2^{4800} MI, wherein objects with TD must be recognized and classified.

The number of classes is very important to set up it about medium as well. Truly, having a few classes, there is a pretty great probability of that 2LP classifies an object randomly correctly. Speaking more strictly, with the number of classes $N_{\mathrm{classes}} \in \mathbb{N} \setminus \{1\}$ this random guessing probability is $N_{\mathrm{classes}}^{-1}$. With overstated number of classes 2LP requires a greater number of neurons in its single hidden layer (SHL), what hampers the fast training process and operation speed of the trained 2LP. Thus may an MI model be the enlarged English alphabet capital letter by $N_{\mathrm{classes}} = 26$ intelligibly. These 26 NDR of classes in GT have some common traits intrinsic to any 60×80 MI (horizontal and vertical lines, diagonals, circles, slants, squares, curves or serpentine lines, etc.). Such MI model ensures additionally that classification results would be allowable to propagate them on other turned objects of the medium format or of a few thousands of their features.

NUMBER OF NEURONS IN PERCEPTRON HIDDEN LAYERS AND MATLAB FUNCTION FOR TRAINING

Number of neurons in 2LP is defined only for its SHL. The input layer has 4800 neurons. The output layer has $N_{\text{classes}} = 26$ neurons. The size of SHL in 2LP for the problem of classifying turned monochrome 60-by-80-images (TM6080I) can be set up to 250 neurons, adequately for GT of 26 NDR of classes in it.

Thus the configuration of 2LP is 4800-250-26 (input layer neurons number — SHL neurons number — output layer neurons number). This 4800-250-26 perceptron (4800-250-26-P) is initialized on MATLAB Neural Network Toolbox simply with the function "feedforwardnet" (the similar function "newff" if for MATLAB versions before R2010) by 1206776 weight and bias values. For training perceptrons there is the backpropagation algorithm, having many methods of its implementation in MATLAB. One of them is the method, where weight and bias values are updated according to gradient descent with adaptive learning rate [9, 10]. It's a MATLAB function "traingda" [11], guaranteeing quick training process passage. Later on, every 4800-250-26-P shall be trained with the function «traingda».

RANGE OF PTSDR

While training a 4800-250-26-P with «traingda», PTSDR is constant. In the statement $r = \dfrac{\sigma_{\text{PD}}}{\sigma_{\text{TD}}}$ both SD may vary, so for PTSDR the assignment $r = \dfrac{\sigma_{\text{PD}}^{\langle\max\rangle}}{\sigma_{\text{TD}}^{\langle\max\rangle}}$ is formally more correct. There $\sigma_{\text{PD}}^{\langle\max\rangle}$ is maximum of SD in PD, added to TM6080I for getting pixel-distorted TM6080I (PDTM6080I), that trains 4800-250-26-P swifter; $\sigma_{\text{TD}}^{\langle\max\rangle}$ is maximum of SD in TM6080I or PDTM6080I. The maximum $\sigma_{\text{TD}}^{\langle\max\rangle}$ must have its upper value, beyond which MI turns so that its turning is impracticable for 2LP classifier, when rotation angle is about $\dfrac{\pi}{8}$ and severer (figure 1). Due to a model of TM6080I, an NDR is turned through the angle

$$\beta(k) = \frac{180}{\pi}\sigma_{\text{TD}}^{\langle k\rangle}\xi(k) \tag{2}$$

in degrees around the center point of MI, where $\xi(k)$ is a value of normal variate (NV) with zero expectation and unit variance (ZEUV), raffled at the k-th stage of TM6080I set formation, $k = \overline{1,\, F}$ and the number $F \in \mathbb{N}$ indicates at smoothness in training a 4800-250-26-P [12]. SD $\sigma_{\text{TD}}^{\langle k\rangle}$ in (2) is

$$\sigma_{\text{TD}}^{\langle k\rangle} = \frac{\sigma_{\text{TD}}^{\langle\max\rangle}}{F}\cdot k \quad \forall\, k = \overline{1,\, F}\,. \tag{3}$$

From the angle (2) by (3) it follows that $\sigma_{\text{TD}}^{\langle\max\rangle} = 0.2$ is enough. In the assignment $r = \dfrac{\sigma_{\text{PD}}^{\langle\max\rangle}}{\sigma_{\text{TD}}^{\langle\max\rangle}}$ this value won't be varied anymore.

For substantiating boundaries $r_{\min}$ and $r_{\max}$ of PTSDR range $[r_{\min};\, r_{\max}]$ it remains to determine the low (least) and upper values for $\sigma_{\text{PD}}^{\langle\max\rangle}$ correspondingly. On $\sigma_{\text{PD}}^{\langle\max\rangle} > 2$ PDTM6080I becomes over-distorted, and on $\sigma_{\text{PD}}^{\langle\max\rangle} < 0.005$ PD is imperceptible. Then $r_{\min} = 0.025$ and $r_{\max} = 10$, enclosing the range $[0.025;\, 10]$ to be run through.

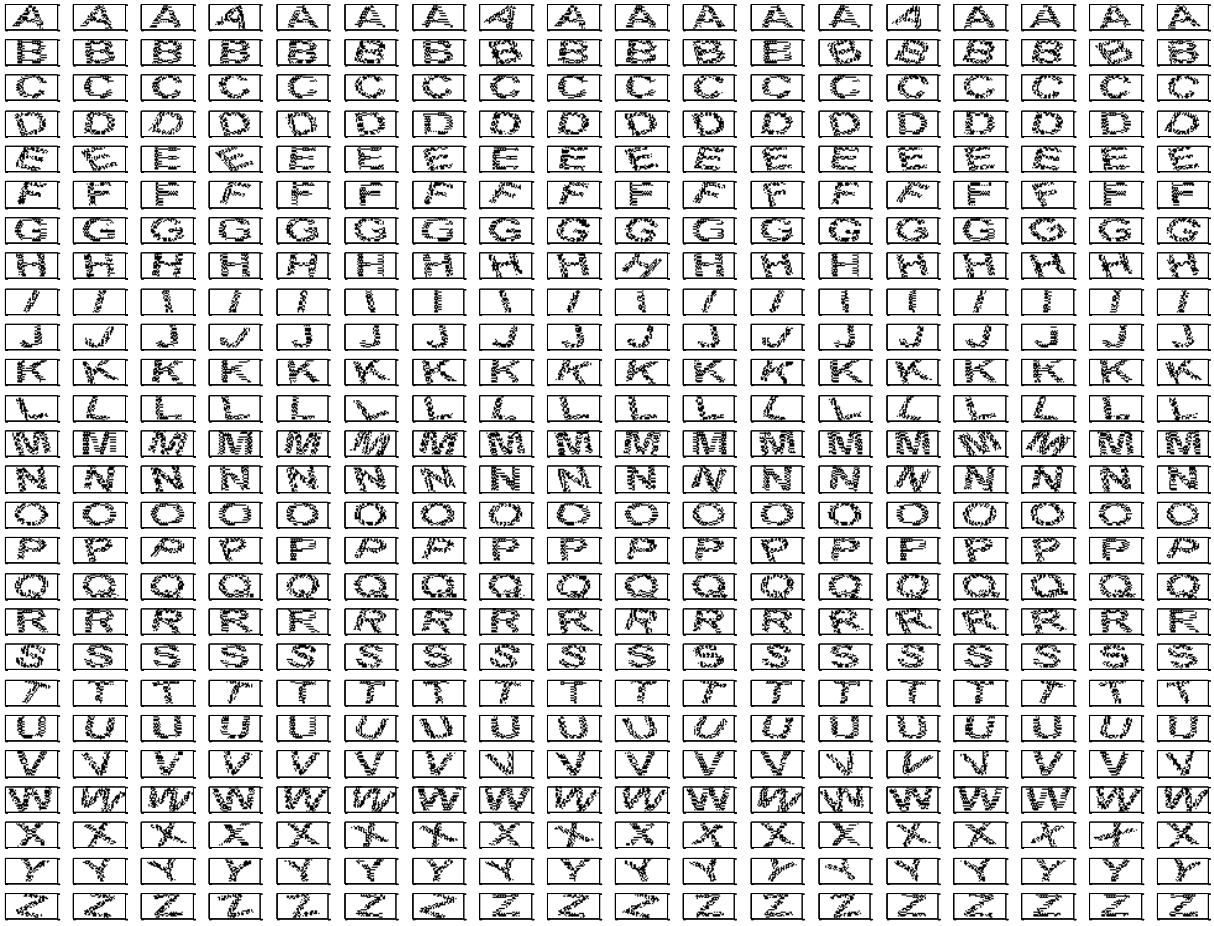

Figure 1. Views of enlarged English alphabet capital letters from GT of 60×80 MI, turned through rotation angles what are too difficult for 2LP to recognize these letters (views from within MATLAB)

RUNNING THROUGH THE RANGE $[0.025;10]$ OF PTSDR

Sampling the range $[0.025;10]$ shouldn't be singly done. There should be a first approximation sampling, after which local minima of the function $p_{\mathrm{EC}}(r)$ are roughly located. The second and, maybe, third approximation samplings are within subranges, constituted by neighborhoods of those minima. These approximations are more precise, needed for locating the optimal PTSDR (1).

Firstly, the range $[0.025;10]$ is sampled with a step, what gives a set of points of PTSDR to train 4800-250-26-P with «traingda» on PDTM6080I, and test the PDTM6080I-trained 4800-250-26-P. In the training process by PDTM6080I the input of 4800-250-26-P is fed with the training set

$$\left\{\mathbf{P}_i^{\langle\mathrm{PDTM6080I}\rangle}\right\}_{i=1}^{C+F}=\left\{\left\{\mathbf{A}\right\}_{l=1}^{C},\left\{\mathbf{A}_{\mathrm{PDTM6080I}}^{\langle k\rangle}\right\}_{k=1}^{F}\right\} \tag{4}$$

of $C\in\mathbb{N}$ replicas of all 26 classes NDR and F matrices of PDTM6080I by targets $\left\{\mathbf{T}_i\right\}_{i=1}^{C+F}=\left\{\mathbf{I}\right\}_{i=1}^{C+F}$ with identity 26×26 matrix $\mathbf{I}$, where 4800×26 matrix $\mathbf{A}$ is formed by concatenating horizontally 4800-length-column-reshaped matrices (4800-LCRM) $\left\{\mathbf{A}_q\right\}_{q=1}^{26}$ by the q-th class NDR as 60×80 MOZ $\mathbf{A}_q=\left(a_{uv}^{\langle q\rangle}\right)_{60\times80}$. At the k-th stage of PDTM6080I set formation, 4800×26 matrix

$$\mathbf{A}_{\mathrm{PDTM6080I}}^{\langle k\rangle}=\mathbf{A}_{\mathrm{TM6080I}}^{\langle k\rangle}+\sigma_{\mathrm{PD}}^{\langle k\rangle}\cdot\boldsymbol{\Xi} \tag{5}$$

by

$$\sigma_{\mathrm{PD}}^{\langle k\rangle}=\frac{\sigma_{\mathrm{PD}}^{\langle\max\rangle}}{F}\cdot k\quad\forall\ k=\overline{1,\ F} \tag{6}$$

and 4800×26 matrix $\boldsymbol{\Xi}$ of values of NV with ZEUV. Concurrently for (5), 4800×26 matrix $\mathbf{A}_{\text{TM6080I}}^{\langle k\rangle}=\left[\bar{\bar{a}}_{jq}(k)\right]_{4800\times 26}$ is formed by concatenating horizontally 4800-LCRM $\left\{\hat{\mathbf{A}}_q(k)\right\}_{q=1}^{26}$ by that q-th TM6080I as the matrix $\hat{\mathbf{A}}_q(k)=\left[\hat{a}_{uv}^{\langle q\rangle}(k)\right]_{60\times 80}$ is the q-th class TD representative. Every single MI as the q-th class NDR $\mathbf{A}_q$ becomes TM6080I

$$\hat{\mathbf{A}}_q(k)=1-\boldsymbol{\rho}\left(1-\mathbf{A}_q,\beta(k),M,S\right) \tag{7}$$

with MATLAB function «imrotate» in the map $\boldsymbol{\rho}$, turning the input MI negative $1-\mathbf{A}_q$ through the angle (2), by the interpolation method handle M and the handle S for specifying the size of the returned negative MI. MI is turned in counterclockwise direction if $\beta(k)>0$, and for $\beta(k)<0$ MI is turned clockwise; for $\beta(k)=0$ MI remains NDR. After all 26 classes NDR have been turned and become TM6080I, the training set (4), formed with (5) — (7), feeds the input of 4800-250-26-P, passing through 4800-250-26-P with targets $\left\{\mathbf{T}_i\right\}_{i=1}^{C+F}=\left\{\mathbf{I}\right\}_{i=1}^{C+F}$ for Q_{pass} times.

PDTM6080I-trained 4800-250-26-P under set $\left\{C,F,Q_{\text{pass}}\right\}=\left\{2,8,10\right\}$ for (4) — (7) is tested with TM6080I at $\sigma_{\text{TD}}\in\left[0;\sigma_{\text{TD}}^{\langle\max\rangle}\right]=[0;0.2]$. If being tested with PDTM6080I, there is $\sigma_{\text{TD}}\in\left[0;\sigma_{\text{TD}}^{\langle\max\rangle}\right]=[0;0.2]$ and $r=5$ as a prime probe, that is $\sigma_{\text{PD}}\in\left[0;\sigma_{\text{PD}}^{\langle\max\rangle}\right]=[0;1]$ by an empiric suggestion $\sigma_{\text{PD}}^{\langle\max\rangle}=1$, where for $\sigma_{\text{TD}}=0$ let it count $r=0$. Generally stating, in a single point of an SD σ by a constant PTSDR r, the testing over 4800-250-26-P with MI of the distortion type $\boldsymbol{\tau}$ produces CEP $p_{\text{EC}}(r,\sigma,\boldsymbol{\tau})$ by $r\in\left[r_{\min};r_{\max}\right]$ at $\sigma\in\boldsymbol{S}$ for $\boldsymbol{\tau}\in\boldsymbol{\Theta}$, where $\boldsymbol{S}$ is a range of SD, $\boldsymbol{\Theta}$ is the set of distortion types. Distortion type $\boldsymbol{\tau}$ can be denoted as linguistic variable (with values “TM6080I”, “PDTM6080I”, etc.), as well as numeric one. For determinacy, let it be numeric variable and $\boldsymbol{\tau}\in\mathbb{N}$. Then that averaged CEP in (1) is a convex combination

$$p_{\text{EC}}(r)=\sum_{\boldsymbol{\tau}\in\boldsymbol{\Theta}}\lambda(\boldsymbol{\tau})\left(\frac{1}{\|\boldsymbol{S}\|}\int_{\sigma\in\boldsymbol{S}}p_{\text{EC}}(r,\sigma,\boldsymbol{\tau})\,d\sigma\right) \tag{8}$$

by

$$\lambda(\boldsymbol{\tau})>0\quad\forall\,\boldsymbol{\tau}\in\boldsymbol{\Theta}\quad\text{and}\quad\sum_{\boldsymbol{\tau}\in\boldsymbol{\Theta}}\lambda(\boldsymbol{\tau})=1, \tag{9}$$

where coefficient $\lambda(\boldsymbol{\tau})$ in (9) for the convex combination (8) designates concernment of the distortion type $\boldsymbol{\tau}$.

Range $[0;0.2]$ is going to be run through with the step 0.02 (when PTSDR is a constant), which lets evaluate the averaged CEP (8)

$$p_{\text{EC}}(r)=\frac{1}{\sigma_{\text{TD}}^{\langle\max\rangle}}\int_{\sigma_{\text{TD}}\in\left[0;\sigma_{\text{TD}}^{\langle\max\rangle}\right]}\frac{p_{\text{EC}}(r,\sigma_{\text{TD}},1)+p_{\text{EC}}(r,\sigma_{\text{TD}},2)}{2}d\sigma_{\text{TD}}=$$

$$=\int_0^{0.2}\frac{p_{\text{EC}}(r,\sigma_{\text{TD}},1)+p_{\text{EC}}(r,\sigma_{\text{TD}},2)}{0.4}d\sigma_{\text{TD}} \tag{10}$$

in those 11 different points of SD in TM6080I and PDTM6080I as

$$p_{\text{EC}}(r)=\sum_{t=0}^{10}\frac{p_{\text{EC}}(r,0.02t,1)+p_{\text{EC}}(r,0.02t,2)}{22}, \tag{11}$$

where the norm $\|\boldsymbol{S}\|$ for summing up (11) has been substituted with cardinality of the sampled subset

$$\{0.02t\}_{t=0}^{10} \subset \left[0;\ \sigma_{\mathrm{TD}}^{\langle\max\rangle}\right] = [0;\ 0.2]$$

and $\boldsymbol{\tau} = 1$ for TM6080I, $\boldsymbol{\tau} = 2$ for PDTM6080I.

As it has been expected (figure 2), the first approximation over the subset

$$\left\{0.025,\ 0.05,\ \{0.1h\}_{h=1}^{20},\ \{2+h\}_{h=1}^{8}\right\} \subset [0.025;\ 10] \tag{12}$$

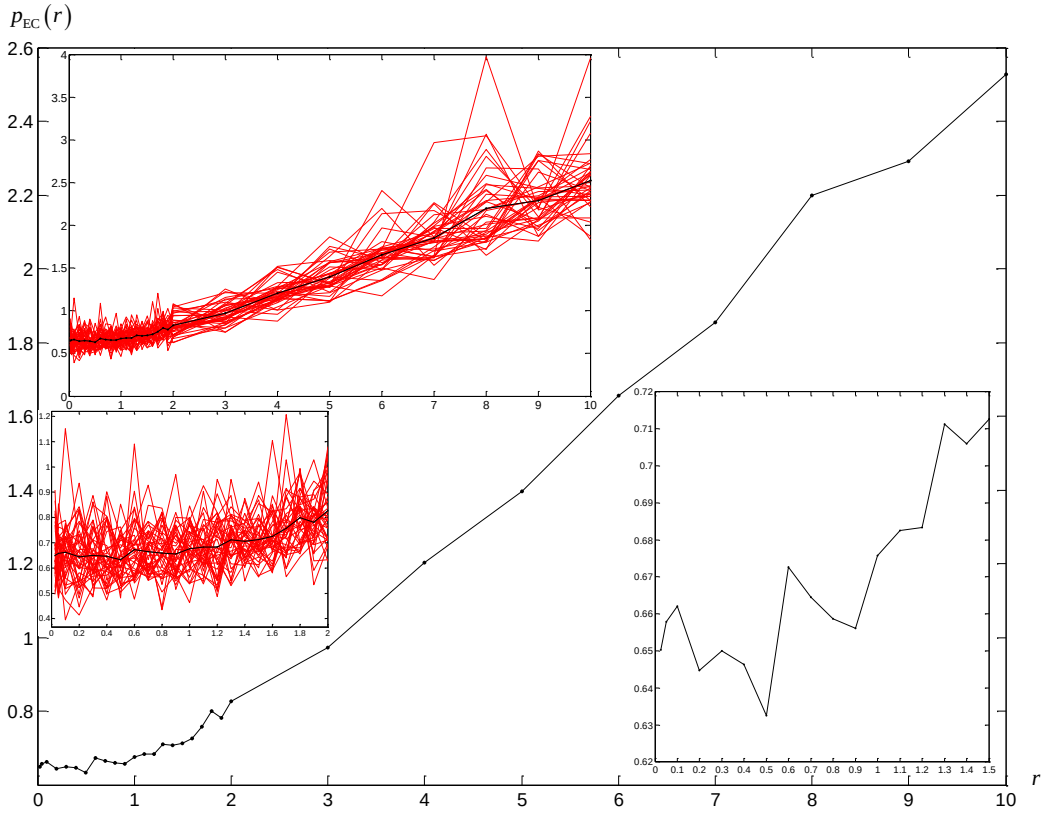

Figure 2. A rough 29-segmented polyline of the function $p_{\mathrm{EC}}(r)$ by (8), derived from 40 series of 400 batch testings of PDTM6080I-trained 4800-250-26-P under set $\{C,\ F,\ Q_{\mathrm{pass}}\} = \{2,\ 8,\ 10\}$ at $\sigma_{\mathrm{TD}}^{\langle\max\rangle} = 0.2$ by $\boldsymbol{\Theta} = \{1\}$ (tested on only TM6080I for faster preliminary results) for each of 30 points (12) of the range $[0.025;\ 10]$

wouldn't have solved the problem (1)

$$r^{*} \in \arg\min_{r\in[0.025;\,10]} p_{\mathrm{EC}}(r). \tag{13}$$

Notwithstanding that PDTM6080I-trained 4800-250-26-P were tested on only TM6080I, figure 2 prompts that the range $[0.025;\,10]$ could be narrowed to the subsegment $[0.025;\,1]$, where $r^{*} \in [0.025;\,1]$ apparently.

The second approximation over the subset

$$\{0.025h\}_{h=1}^{40} \subset [0.025;\ 1] \tag{14}$$

gives fuzzy imagination about locations of the function $p_{\mathrm{EC}}(r)$ minima (figure 3). And twice greater number of batch testings didn't help. Nevertheless, this 0.025-sampling lets feel and see that $r^{*} \in [0.025;\ 0.6]$, needing though the third approximation.

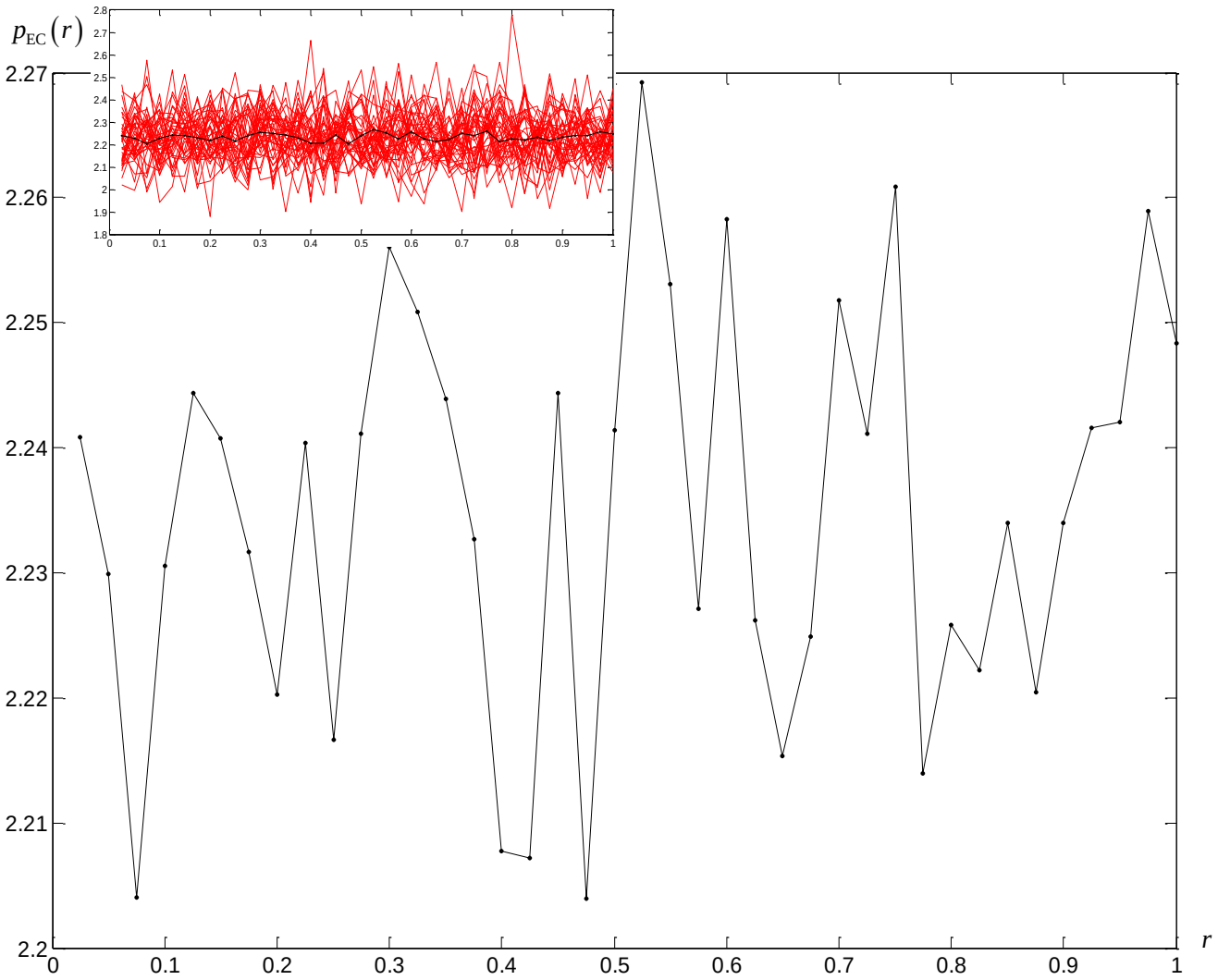

Figure 3. Localized 39-segmented polyline of the function $p_{EC}(r)$ by (11) on $\sigma_{TD} \in [0; 0.2]$ and $\sigma_{PD} \in [0; 1]$, derived from 40 series of 800 batch testings of PDTM6080I-trained 4800-250-26-P under set $\{C, F, Q_{pass}\} = \{2, 8, 10\}$ at $\sigma_{TD}^{\langle max \rangle} = 0.2$ for each of 40 points (14) of the subrange $[0.025; 1] \subset [0.025; 10]$

Fulfilling still 0.025-sampling again (figure 4), the third approximation over the subset

$$\{0.025h\}_{h=1}^{24} \subset [0.025; 0.6] \tag{15}$$

roughly confirms the fuzzy locations of the function $p_{EC}(r)$ minima about PTSDR points $r \in [0.025; 0.075]$ and $r \in [0.425; 0.525]$, what can be seen also previously in figure 3. However, after having superposed 100 23-segmented polylines of the function $p_{EC}(r)$ from figure 4 and 40 polylines from figure 3 (thanking to that they all 140 were obtained on 0.025-sampling), those fuzzy minima become fuzzier (figure 5) in the averaged 23-segmented polyline of the function $p_{EC}(r)$. Moreover, there is a suspicion on that the minimum point (13) could be about either PTSDR $r = 0.075$ or $r = 0.175$, whereas PTSDR $r \in (0.3; 0.4)$ doesn't seem to be the problem (13) solution. Therefore the testings must be prolonged as the following. Another 200 series of 200 batch testings (what appears to be sufficient to understand classification properties of an PDTM6080I-trained 4800-250-26-P) of PDTM6080I-trained 4800-250-26-P over the subset (15) are to be carried out and analyzed. If the minimum point (13) can be tracked then the testings are stopped, else the extra 100 series are carried out and analyzed until the minimum point (13) is tracked.

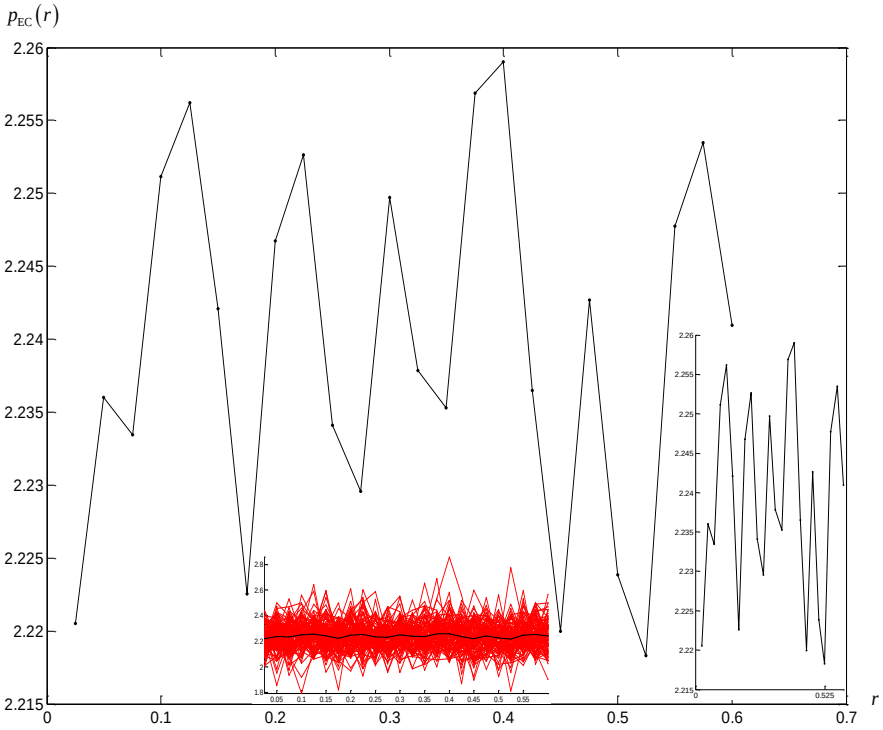

Figure 4. Localized 23-segmented polyline of the function $p_{EC}(r)$ by (11) on $\sigma_{TD} \in [0;\ 0.2]$ and $\sigma_{PD} \in [0;\ 1]$, derived from 100 series of 800 batch testings of PDTM6080I-trained 4800-250-26-P under set $\{C,\ F,\ Q_{pass}\} = \{2,\ 8,\ 10\}$ at $\sigma_{TD}^{\langle max \rangle} = 0.2$ for each of 24 points (15) of the subrange $[0.025;\ 0.6] \subset [0.025;\ 1]$

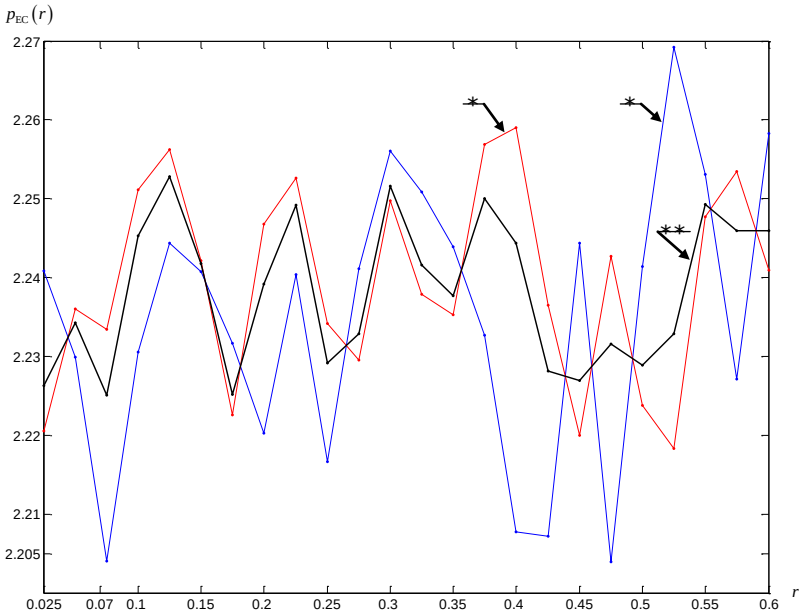

Figure 5. The averaged 23-segmented polyline (***) of the function $p_{EC}(r)$ over 140 polylines, plotted on 24 points (15) of the subrange $[0.025;\ 0.6] \subset [0.025;\ 1]$ (the average of polylines in figures 3 and 4 on this subrange), where polylines (*) and (**) relate to figure 3 and figure 4 correspondingly

Regretfully, but those another 200 series of 200 batch testings of PDTM6080I-trained 4800-250-26-P over the subset (15) nowise augment (figure 6) the 140-polylines-averaged result in figure 5: there are no confirmations of the function $p_{EC}(r)$ minima fuzzy locations, nor superiority of a subsegment of the subrange

$$[0.025;\ 0.6] \subset [0.025;\ 1] \subset [0.025;\ 10]$$

in enclosing the global minimum point. The averaged 23-segmented polyline of the function $p_{\mathrm{EC}}(r)$ over 340 polylines on 24 points (15) strengthens this disillusionment (figure 7). Here, although the point $r = 0.475$ could have been ventured to accept it as the problem (13) solution, it as well could be refuted. The matter is that in localizing 23-segmented polyline of the function $p_{\mathrm{EC}}(r)$ by averaging over a subset of those used previously 340 polylines, the polyline shape in figure 7 fails (figure 8). A subset of indices of the used 340 polylines, indexed as $G = \{g\}_{g=1}^{340}$, is locally determined as a subsample from the set

$$G_{\boldsymbol{\pi}} = \boldsymbol{\pi}(G) = \{j_g\}_{g=1}^{340}, \tag{16}$$

where $\boldsymbol{\pi}(G)$ is a random permutation of the integers from the set $G = \{g\}_{g=1}^{340}$. Then for the number $J_{\mathrm{subsample}} \in \{\overline{1, 10}\}$ of 34-multiple extractions from the set of 340 polylines, the polylines with the set of their indices

$$G_{\boldsymbol{\pi}}^{\langle J_{\mathrm{subsample}} \rangle} = \{j_g\}_{g=1}^{34 \cdot J_{\mathrm{subsample}}} \subset G_{\boldsymbol{\pi}} \tag{17}$$

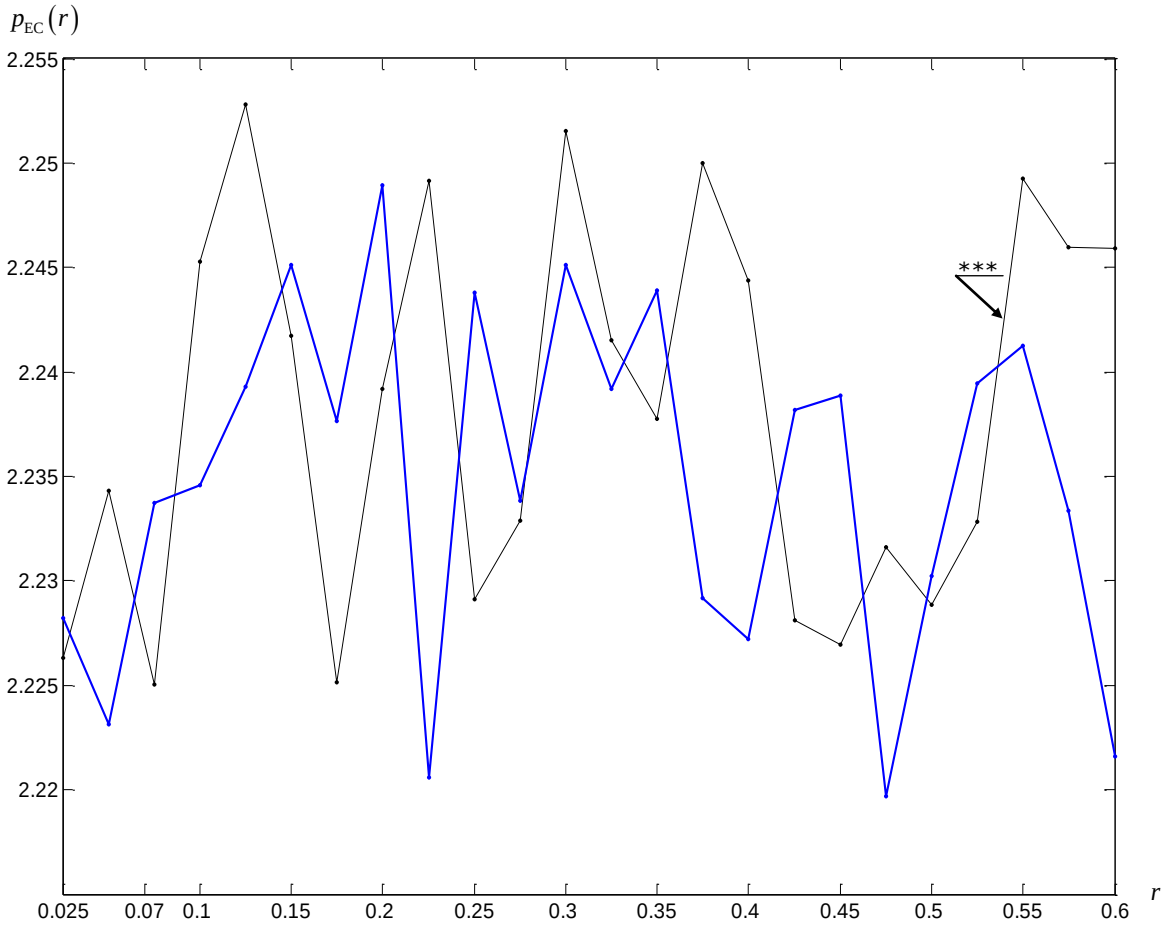

Figure 6. Localized 23-segmented polyline of the function $p_{\mathrm{EC}}(r)$ by (11) on $\sigma_{\mathrm{TD}} \in [0;\ 0.2]$ and $\sigma_{\mathrm{PD}} \in [0;\ 1]$, derived from 200 series of 200 batch testings of PDTM6080I-trained 4800-250-26-P under set $\{C,\ F,\ Q_{\mathrm{pass}}\} = \{2,\ 8,\ 10\}$ at $\sigma_{\mathrm{TD}}^{\langle \max \rangle} = 0.2$ for each of 24 points (15) of the subrange $[0.025;\ 0.6] \subset [0.025;\ 1]$, where polyline (***) relates to fig. 5

are averaged. The greater the number $J_{\mathrm{subsample}}$ is the more reliable average is expected. And if to extract a lesser cardinality subset from those 340 polylines, the shape of the 23-segmented polyline

of the function $p_{\mathrm{EC}}(r)$ becomes stochastic (figure 9), without any resemblances to figure 7. Ultimately, the case $J_{\mathrm{subsample}}=10$ will give a random permutation

$$G_{\boldsymbol{\pi}}^{\langle 10\rangle}=\left\{j_g\right\}_{g=1}^{340}=G_{\boldsymbol{\pi}}$$

over all 340 polylines, whose average is that 23-segmented polyline of the function $p_{\mathrm{EC}}(r)$ in figure 7. Commonly, the maximal number $J_{\mathrm{subsample}}$ shouldn't be necessarily multiple to the total number of polylines, and the maximal number $J_{\mathrm{subsample}}$ cannot be greater than the total number of polylines. Here the number $J_{\mathrm{subsample}}\in\left\{\overline{1,10}\right\}$ is taken purely for convenience in ascertaining the rate of stochasticity in those 340 polylines, whose average is in figure 7. This rate is viewed (or measured) against the averaged polyline in figure 7. The spoken stochasticity is very deep for either $J_{\mathrm{subsample}}=\overline{1,8}$ or $J_{\mathrm{subsample}}=\overline{1,6}$. Those averaged polylines after realizations of extraction resemble stochastic functions regardless.

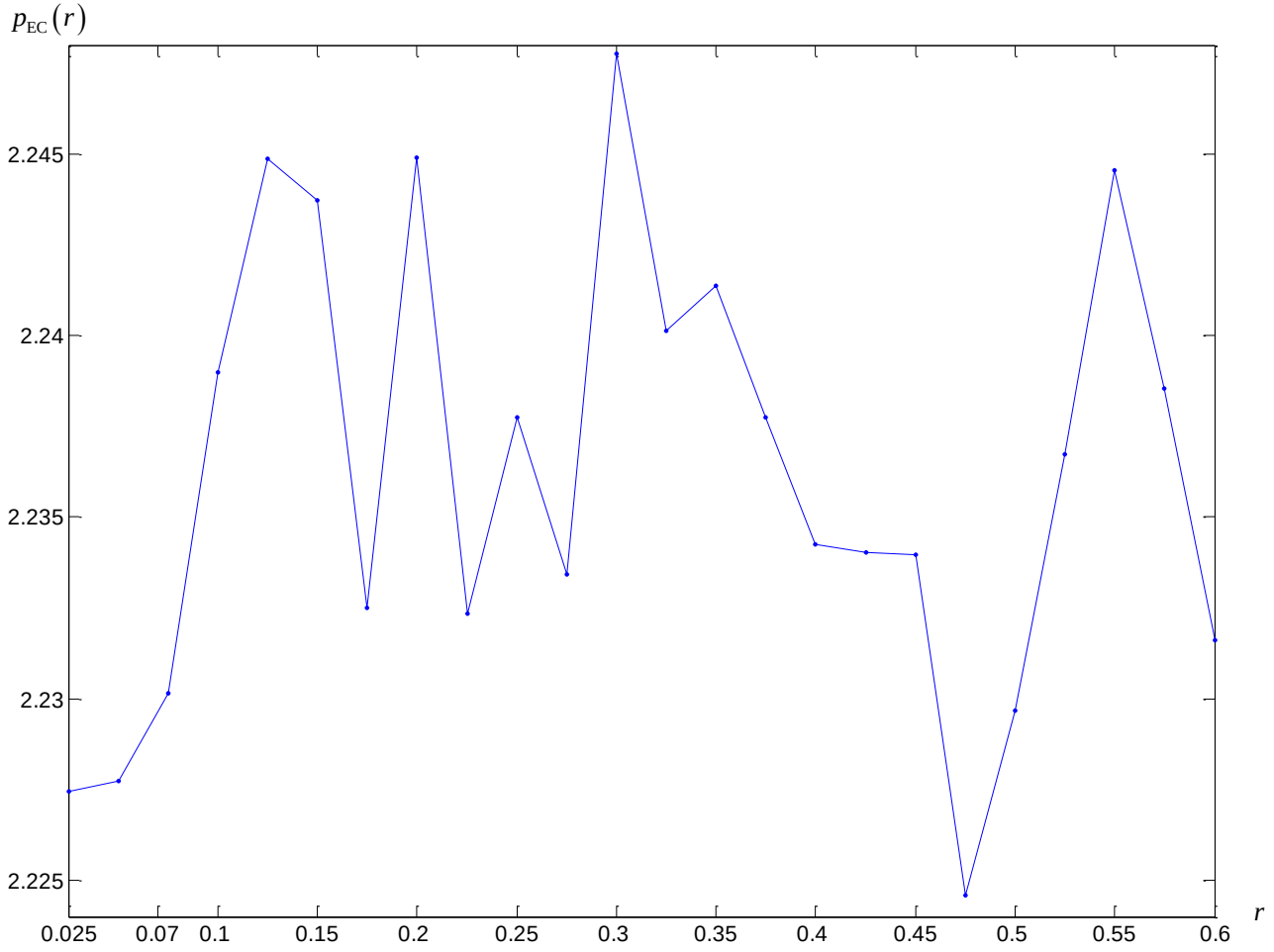

Figure 7. Localized 23-segmented polyline of the function $p_{\mathrm{EC}}(r)$ by (11) on $\sigma_{\mathrm{TD}}\in[0;\,0.2]$ and $\sigma_{\mathrm{PD}}\in[0;\,1]$ as the average of 340 polylines on 24 points (15), drawn from two averaged 23-segmented polylines in figure 6

However, this deep stochasticity may be provoked with the current conformation of training and testing, where $Q_{\mathrm{pass}}=10$ and $\sigma_{\mathrm{TD}}\in[0;\,0.2]$ and $\sigma_{\mathrm{PD}}\in[0;1]$. Yes, number of passes is too low for expecting stable estimation. Another side is that PTSDR in testing is high, $r=5$, what could additionally randomize the output results of estimations. Somehow or other, there ought to be a try to determine a variant of minimum (13) even for this "stochastic" conformation of training and testing.

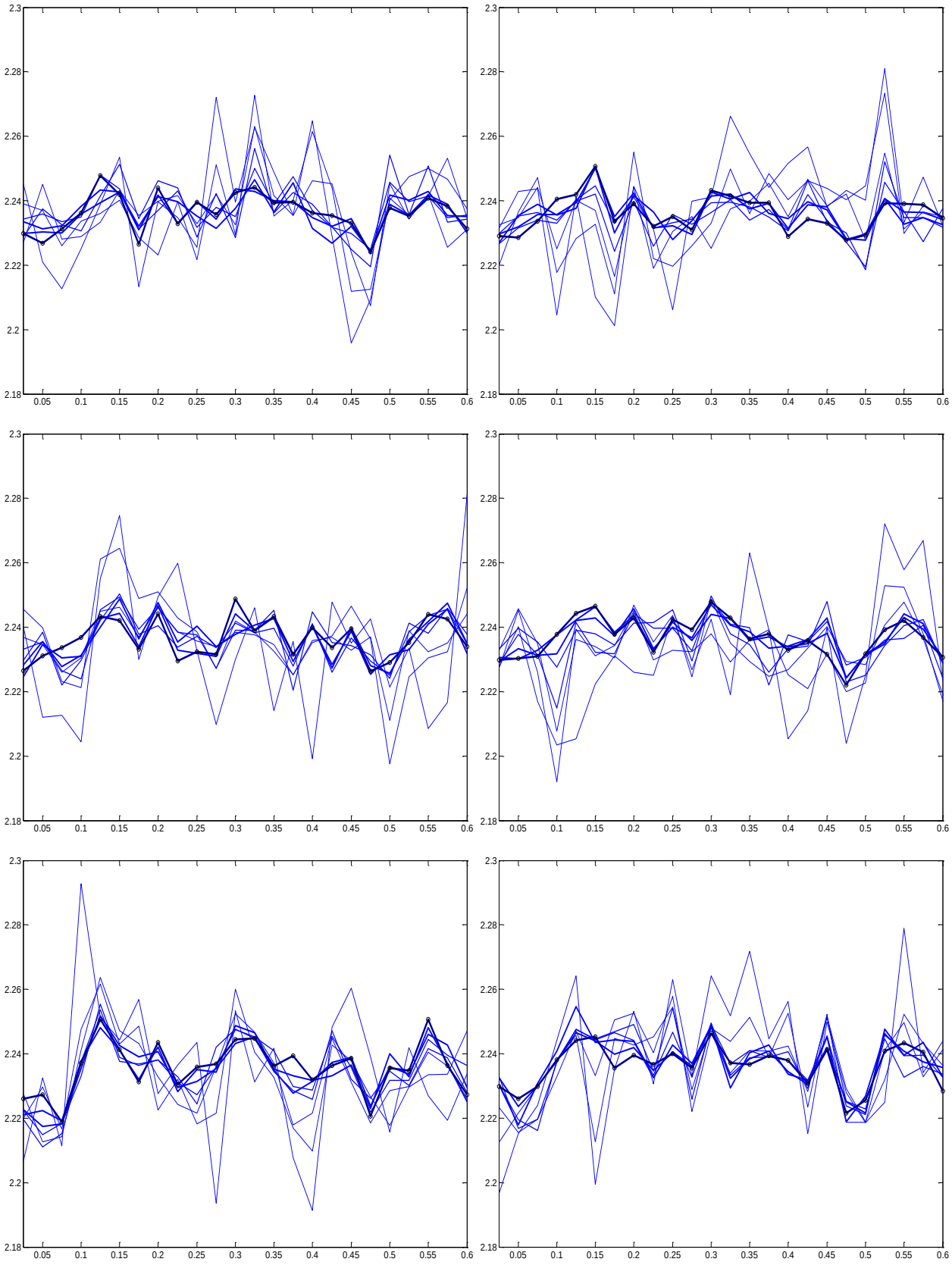

Figure 8. Six realizations of extraction of subsets of the 340 polylines' indices set $G = \{g\}_{g=1}^{340}$ by (16) and (17) for $J_{\text{subsample}} = \overline{1,\ 8}$, where the thicker polyline corresponds to a greater number $J_{\text{subsample}}$ and the resulting circle-dashed polyline is averaged over $\left|G_{\pi}^{\langle 8\rangle}\right| = \left|\{j_g\}_{g=1}^{272}\right| = 272$ polylines; here the point $r = 0.475$ is five times spotted as the global minimum of $p_{\text{EC}}(r)$

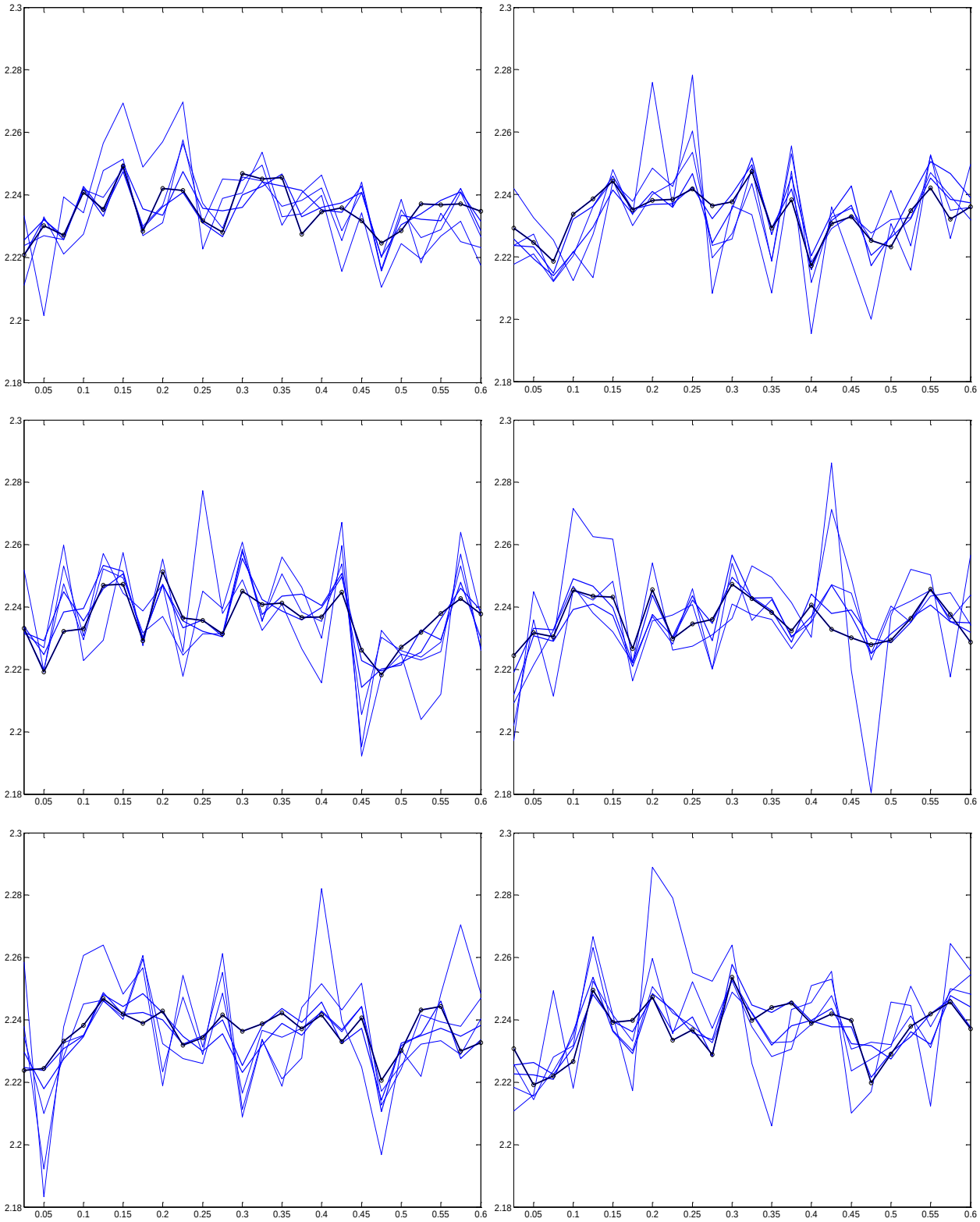

Figure 9. Six realizations of extraction of subsets of the 340 polylines' indices set $G = \{g\}_{g=1}^{340}$ by (16) and (17) for $J_{\text{subsample}} = \overline{1,\ 6}$, where the thicker polyline corresponds to a greater number $J_{\text{subsample}}$ and the resulting circle-dashed polyline is averaged over $\left|G_{\pi}^{\langle 6\rangle}\right| = \left|\left\{j_g\right\}_{g=1}^{204}\right| = 204$ polylines, appearing more stochastic in comparison to figure 7 and figure 8; here the point $r = 0.475$ is two times spotted as the global minimum of $p_{\text{EC}}(r)$

The stochastics in figures 8 and 9, not weakening the stochastics in figures 3 — 6, and the resulting polyline in figure 7 might be evidence of that the problem (13) solution is in the segment $[0.025;\ 0.5]$ of PTSDR. And it is impossible to track the single minimum point (13) with studying

extra series of PDTM6080I-trained 4800-250-26-P. Nevertheless, the single minimum point (13) within the segment $[0.025;\ 0.5]$ of PTSDR might have been found by applying the criterion of the time spent for the training process. Figure 10, showing the normed polyline from figure 7 and the normed averaged time spent for the training process on the same axes, testifies that there is a global minimum variant of $r^* = 0.5$.

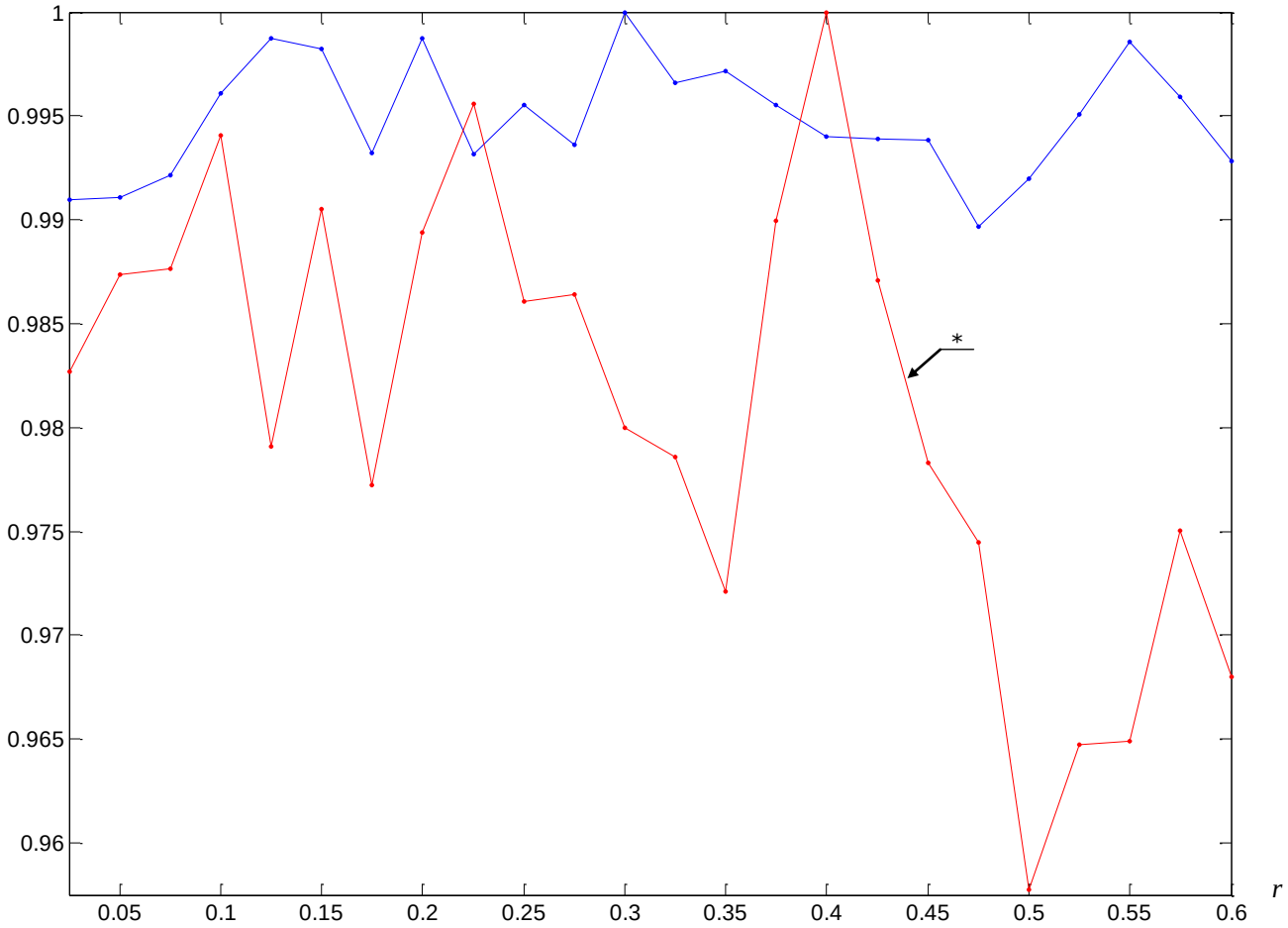

Figure 10. Normed polyline of the function $p_{\text{EC}}(r)$ from figure 7 and the normed averaged time (*) spent for the training process of 200 PDTM6080I-trained 4800-250-26-P in figure 6

Consequently, PTSDR $r = 0.5$, suspected to be the problem (13) solution since figure 2 with 0.025 -deviation by figures 3, 4, 6, 7, could have been out for verifying whether CEP $p_{\text{EC}}(0.5)$ is minimal over TM6080I and PDTM6080I. However, to decrease CEP deeper one would have to increase the number Q_{pass}, what is expected to shift such unstable solution. Thus all the stated within a prime probe is just corroborating the fact that the problem (13) solution is very hard to get it fixed under low number Q_{pass} and too high $\sigma_{\text{PD}}^{\langle\max\rangle}$ in testing with PDTM6080I, what stimulates stochastic features much.

Well, next probe is going to be under $Q_{\text{pass}} = 25$ and PDTM6080I-trained 4800-250-26-P will be tested with PDTM6080I on $\sigma_{\text{TD}} \in [0;\ 0.2]$ and $\sigma_{\text{PD}} \in [0;\ 0.2]$. The first approximation over the subset

$$\left\{0.025,\ 0.05,\ \{0.1h\}_{h=1}^{20}\right\} \subset [0.025;\ 2] \tag{18}$$

hints at that the global minimum is shifted to the utmost left (figure 11). At $r = 0.05$ there is a local maximum nearest to the global minimum. Of course, it is statistically not so reliable. Nonetheless the shape in figure 11 resembles much the shape in figure 2. Clearly, there is no global minimum at

$r>1$. Local minimum at $r=1.1$ is just exposition of statistical instability. Local minimum at $r=0.7$ is of the same lack of statistics (it took a long period to obtain those 160 series, not speaking about greater number of series, what would take much longer). Local maximum at $r=0.05$ is too unexpected. The polyline is monotonously increasing by $r\in[0.1;\,0.6]$ and $r\in[0.7;\,1]$, and is monotonously non-decreasing by $r\in[1.1;\,2]$. Not viewing the polyline in detail, it seems to be increasing exponentially on average.

Without needless argumentation, the second approximation follows. For this, the subset (18) is narrowed to the subset

$$\{0.025h\}_{h=1}^{32}\subset[0.025;\,0.8] \tag{19}$$

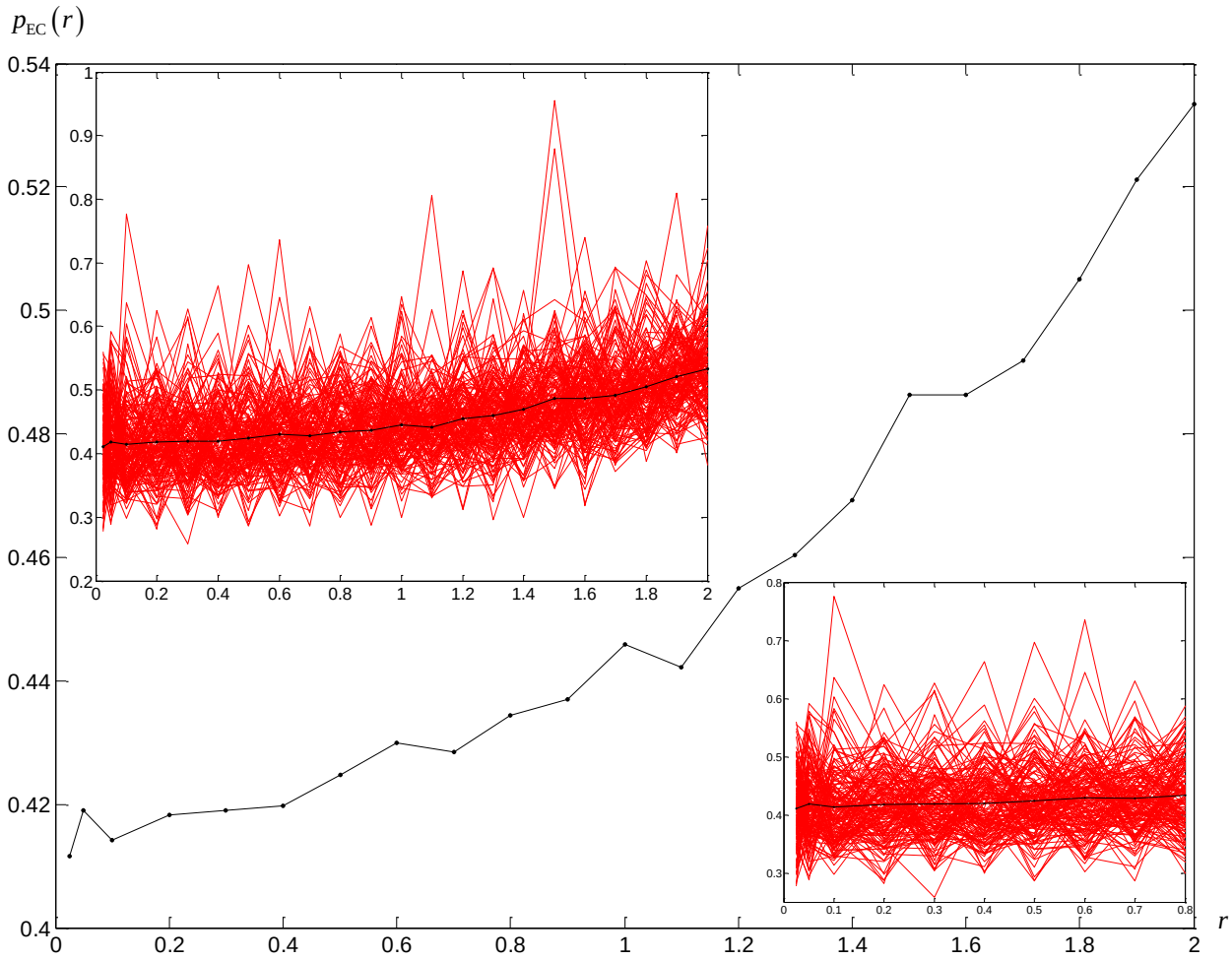

Figure 11. A 21-segmented polyline of the function $p_{EC}(r)$ by (11) on $\sigma_{TD}\in[0;\,0.2]$ and $\sigma_{PD}\in[0;\,0.2]$, derived from 160 series of 400 batch testings of PDTM6080I-trained 4800-250-26-P under set $\{C,\,F,\,Q_{pass}\}=\{2,\,8,\,25\}$ at $\sigma_{TD}^{\langle max\rangle}=0.2$ for each of 22 points (18) of the subrange $[0.025;\,2]\subset[0.025;\,10]$

enclosing the global minimum point certainly. Figure 12 shows the polyline, plotted after a really huge number of series of training and testing. But stochasticity is still present — there are no long monotonous parts along the polyline. This polyline fluctuates heavily, and the increase on average is hardly examined. There is only three-segmented monotonous part, where by $r\in[0.425;\,0.5]$ the polyline is increasing. The global minimum over the subset (19) has appeared at $r=0.175$, although the left-ended local minimum is closer than others.

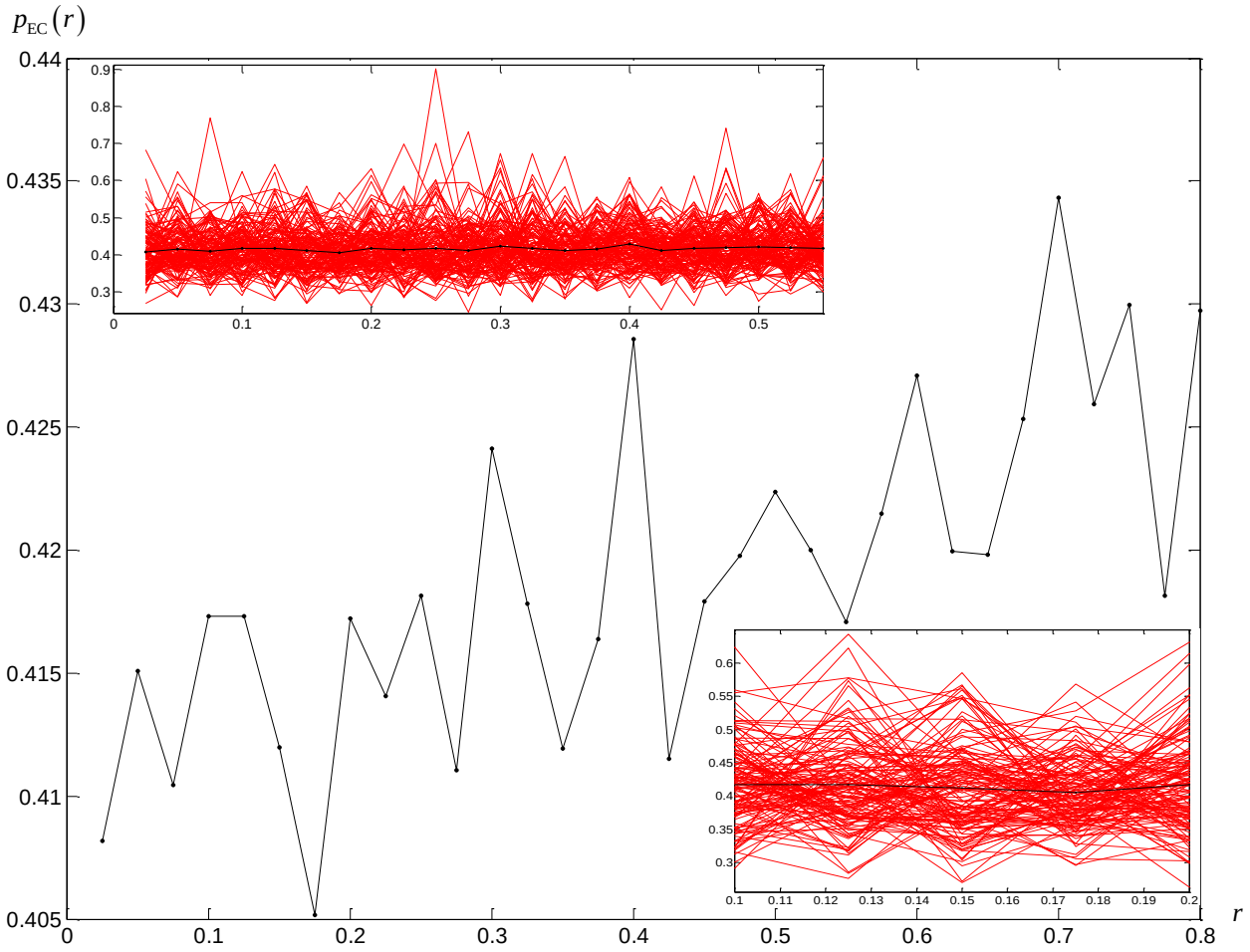

Figure 12. Localized 31-segmented polyline of the function $p_{EC}(r)$ by (11) on $\sigma_{TD} \in [0;\ 0.2]$ and $\sigma_{PD} \in [0;\ 0.2]$, derived from 140 series of 400 batch testings of PDTM6080I-trained 4800-250-26-P under set $\{C,\ F,\ Q_{pass}\} = \{2,\ 8,\ 25\}$ at $\sigma_{TD}^{\langle max \rangle} = 0.2$ for each of 32 points (19) of the subrange $[0.025;\ 0.8] \subset [0.025;\ 2]$

Therefore, in this conformation of training and testing, where $Q_{pass} = 25$ and $\sigma_{TD} \in [0;\ 0.2]$ and $\sigma_{PD} \in [0;\ 0.2]$, the second variant of global minimum is $r^* = 0.175$. Note that PTSDR in testing has been five times decreased, and the global minimum variant has slid to the left from $r^* = 0.5$ to $r^* = 0.175$, what is slightly anticipated. But the polyline in figure 12 doesn't look less stochastic than polylines in figures 4 — 10. It's an evidence of that PTSDR for testing doesn't influence much on stochasticity. It influences rather on the problem (13) solution.

Another peculiarity is that the left-endpoint shift of the minimum r^* isn't desirable. It might be regarding to that PTSDR in testing has been over-decreased. Thus, leaving $Q_{pass} = 25$ and $\sigma_{TD} \in [0;\ 0.2]$, the value $\sigma_{PD}^{\langle max \rangle}$ in testing with PDTM6080I is going to be increased. This increment is hoped to get the third variant of the minimum r^*, which should be closer to the second one $r^* = 0.175$ by figure 12. Value of the increment is surely significant. But it is put simply to be twice greater than the previous PTSDR in testing.

Well, the third probe is going to be under $Q_{pass} = 25$ and PDTM6080I-trained 4800-250-26-P will be tested with PDTM6080I on $\sigma_{TD} \in [0;\ 0.2]$ and $\sigma_{PD} \in [0;\ 0.4]$. The first approximation over the subset (18) is in figure 13. The averaged CEP increased and no wonder — PD is twice higher. Resemblance of polylines in figures 11 and 13 is out of dispute. Again we see that the global minimum is shifted to the utmost left. A strange local maximum nearest to the global minimum is at

$r = 0.1$ now. Moving along PTSDR axis to the right, there are only two local maximums — at $r = 0.4$ and $r = 0.6$, although the local maximum at $r = 0.4$ is more apparent and the local maximum at $r = 0.6$ looks like just a statistical splash.

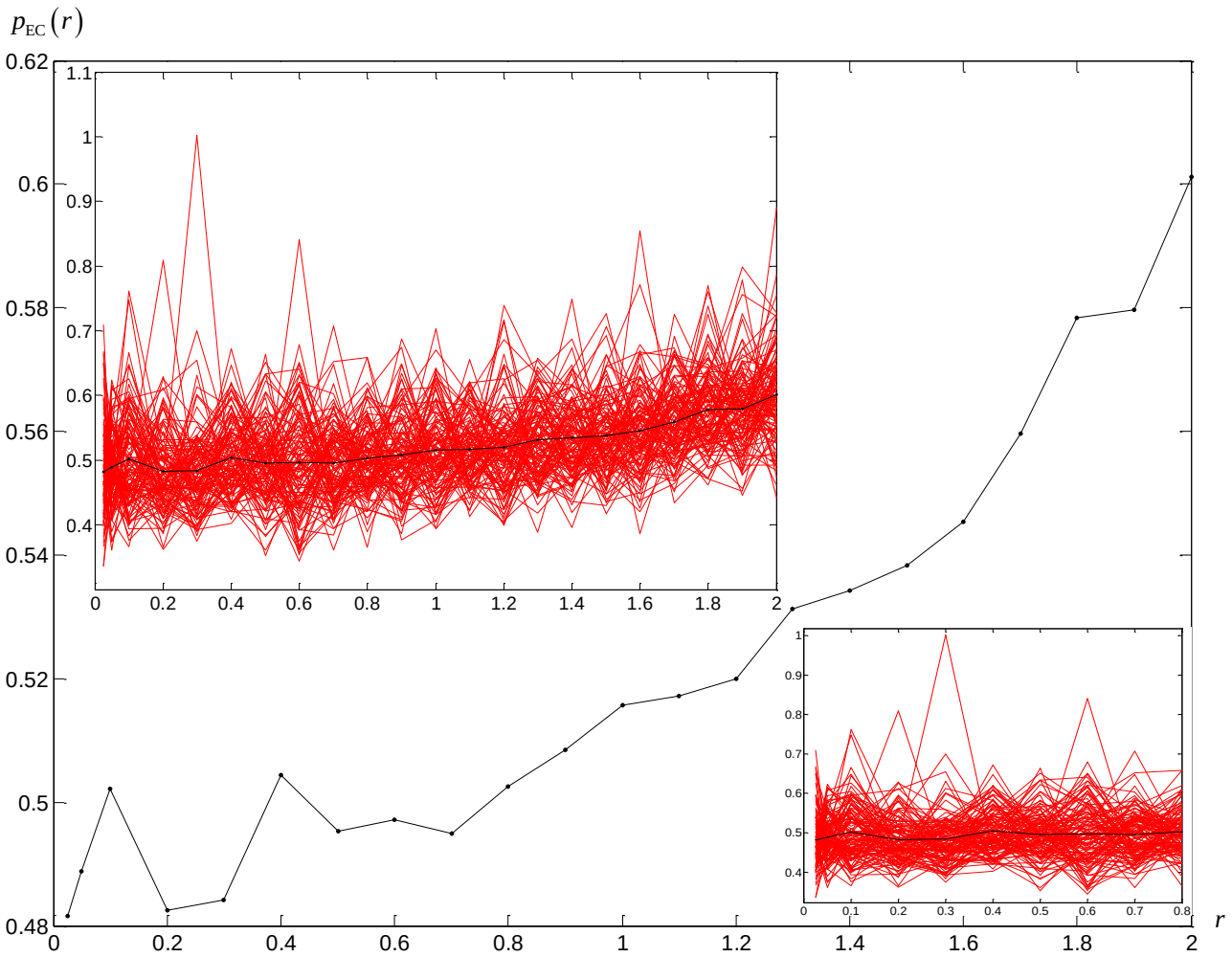

Figure 13. A 21-segmented polyline of the function $p_{EC}(r)$ by (11) on $\sigma_{TD} \in [0;\ 0.2]$ and $\sigma_{PD} \in [0;\ 0.4]$, derived from 120 series of 400 batch testings of PDTM6080I-trained 4800-250-26-P under set $\{C,\ F,\ Q_{pass}\} = \{2,\ 8,\ 25\}$ at $\sigma_{TD}^{\langle max \rangle} = 0.2$ for each of 22 points (18) of the subrange $[0.025;\ 2] \subset [0.025;\ 10]$

It feels like the problem (13) solution is $r^* < 1$. It's less sure about the problem (13) solution is $r^* < 0.8$, but the polyline in figure 13 is monotonously increasing off PTSDR $r = 0.7$ along PTSDR axis to the right. Without needless argumentation, the second approximation follows. For this, the subset (18) is narrowed to the subset (19) enclosing the global minimum point certainly. Figure 14 shows the polyline, plotted after a really huge number of series of training and testing. This polyline stochasticity is similar to stochasticity in figure 12 — there are no long monotonous parts along the polyline. The polyline in figure 14 fluctuates heavily, and the increase on average is hardly examined, especially for $r < 0.6$. Strangeness of local maximum at $r = 0.05$ is explained with the same lack of statistics. The rest seven local maximums for $r < 0.8$ have the same origin. Nine local minimums within the subset (19) have similarly stochastic locations. Now the global minimum over the subset (19) has appeared at $r = 0.15$, and the closest local minimum by the value of CEP is reached at $r = 0.5$, although local minimums at $r = 0.025$ and $r = 0.25$ and $r = 0.3$ give almost the same values of CEP.

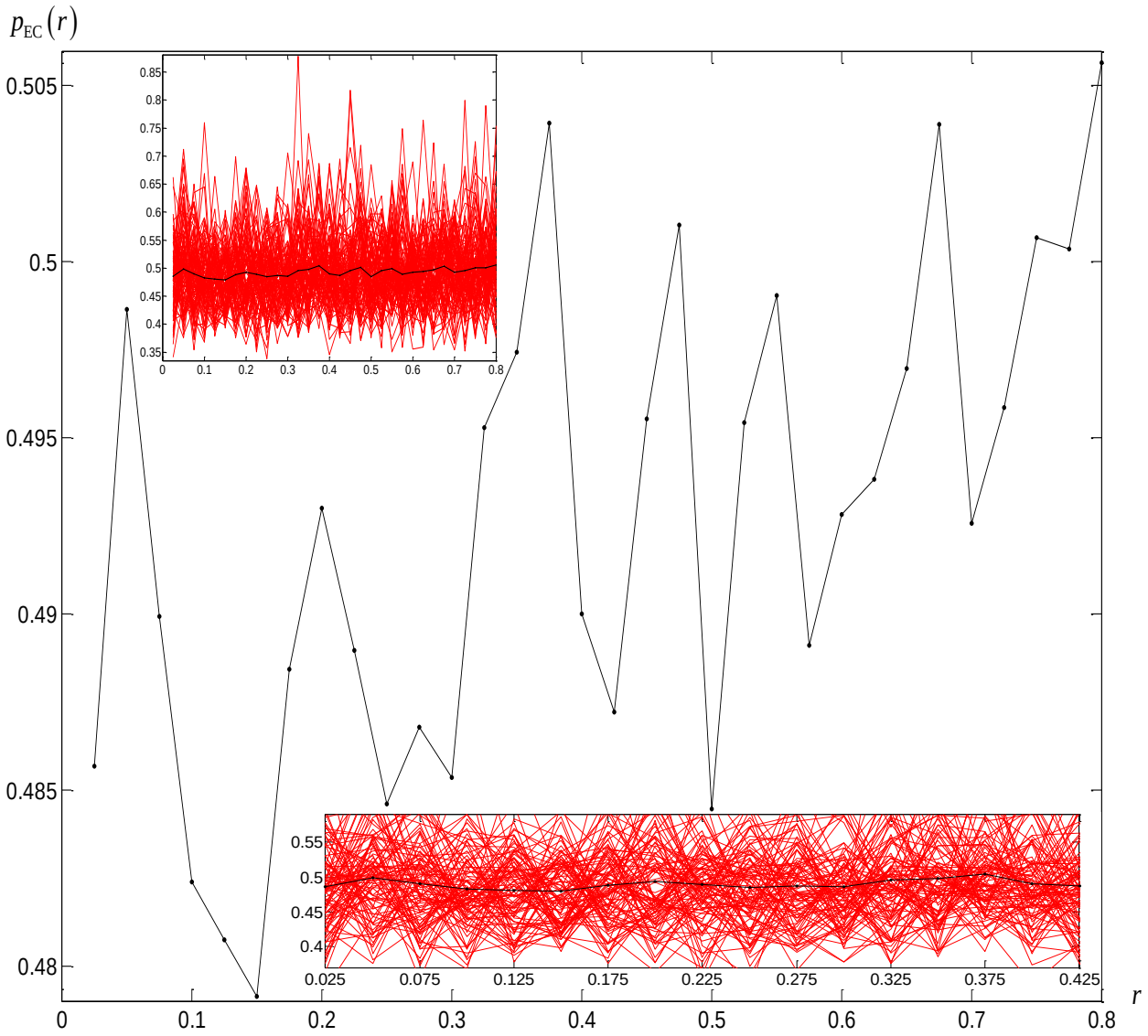

Figure 14. Localized 31-segmented polyline of the function $p_{EC}(r)$ by (11) on $\sigma_{TD} \in [0;\ 0.2]$ and $\sigma_{PD} \in [0;\ 0.4]$, derived from 150 series of 400 batch testings of PDTM6080I-trained 4800-250-26-P under set $\{C,\ F,\ Q_{pass}\} = \{2,\ 8,\ 25\}$ at $\sigma_{TD}^{\langle max \rangle} = 0.2$ for each of 32 points (19) of the subrange $[0.025;\ 0.8] \subset [0.025;\ 2]$

Obviously, the second and third variants of the problem (13) solution are pretty close, differing only in 0.025 units of PTSDR or in the single step after 0.025-sampling. An important current question is whether the third variant $r^* = 0.15$ maintain one's location if to re-localize the polyline of the function $p_{EC}(r)$ by (11) over shorter interval. Thus the third approximation follows. For this, the subset (19) is narrowed to the subset

$$\{0.025h\}_{h=1}^{16} \subset [0.025;\ 0.4] \tag{20}$$

anticipating that the non-sampled segment

$$[0.025;\ 0.4] \subset [0.025;\ 0.8] \subset [0.025;\ 2] \tag{21}$$

encloses the global minimum point certainly. Figure 15 shows the result of huger number of series of training and testing, though 150 series of 400 batch testings were taken from Figure 14.

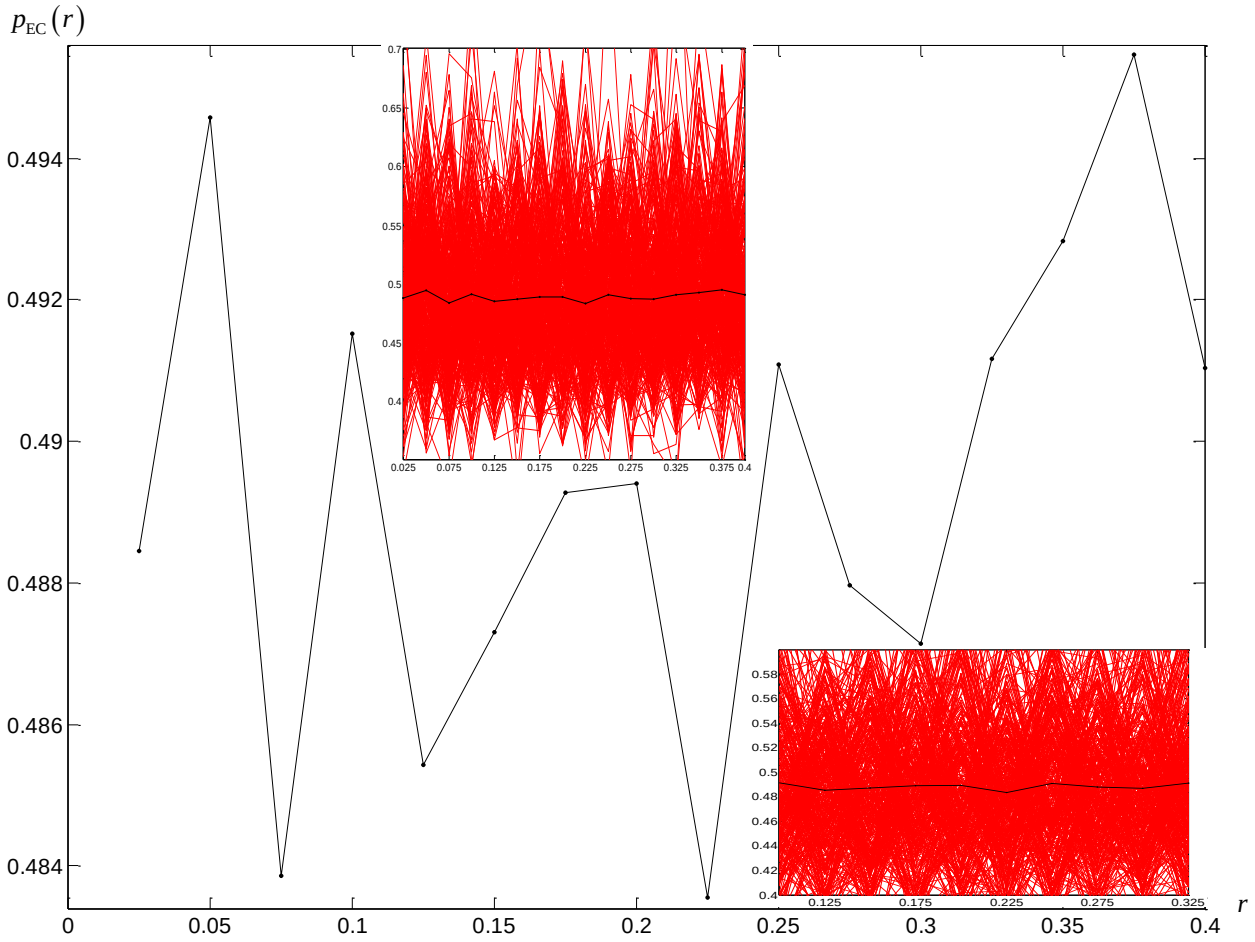

Figure 15. Localized 15-segmented polyline of the function $p_{EC}(r)$ by (11) on $\sigma_{TD} \in [0;\ 0.2]$ and $\sigma_{PD} \in [0;\ 0.4]$, derived from 370 series of 400 batch testings of PDTM6080I-trained 4800-250-26-P under set $\{C,\ F,\ Q_{pass}\} = \{2,\ 8,\ 25\}$ at $\sigma_{TD}^{\langle max \rangle} = 0.2$ for each of 16 points (20) of the subrange (21)

Honestly, the polyline view is once again telling about great level of stochasticity in estimating the function $p_{EC}(r)$ by (11). The third variant of the problem (13) solution $r^* = 0.15$ shifted to the right, and $r^* = 0.225$ now. Besides, the local minimum at $r = 0.075$ gives almost the same value of CEP. Local maximums at $r = 0.05$ and at $r = 0.1$ are particularly surprising. In this situation, the most appropriate conclusion is that $r^* \in (0.075;\ 0.3)$, where one can really see cavity of the polyline in figure 15 by $r \in (0.075;\ 0.3)$ or, in terms of 0.025-sampling, by

$$r \in \{0.025h\}_{h=3}^{12} \subset \{0.025h\}_{h=1}^{16} \subset [0.025;\ 0.4].$$

The similar cavity is overviewed in figure 14 by $r \in (0.1;\ 0.3)$ or, in terms of 0.025-sampling, by

$$r \in \{0.025h\}_{h=4}^{12} \subset \{0.025h\}_{h=1}^{16} \subset [0.025;\ 0.4].$$

Roughly watching, the cavity can be overviewed also in figure 13 by $r \in (0.2;\ 0.3)$ or, including neighboring local maximums, by $r \in (0.1;\ 0.4)$. Contrariwise, polyline in figure 12 doesn't have any visible cavities. Notwithstanding the obvious fuzziness in estimating the function $p_{EC}(r)$ by (11) over shorter intervals with those cavities, the final decision about the problem (13) solution is going to be made easily. By applying the criterion of the time spent for the training process (figure 10), it is rather to have higher PTSDR, which may slightly accelerate training. Finally, the single minimum point (13) by visualization in figure 15 is $r^* = 0.225$.

THE AVERAGED CEP MINIMIZATION DUE TO THE PROBLEM (13)

Statistics of the PDTM6080I-trained 4800-250-26-P under sets

$$\left\{C,\, F,\, Q_{\text{pass}},\, \sigma_{\text{TD}}^{\langle\max\rangle}\right\} = \{2,\, 8,\, 10,\, 0.2\}$$

and

$$\left\{C,\, F,\, Q_{\text{pass}},\, \sigma_{\text{TD}}^{\langle\max\rangle}\right\} = \{2,\, 8,\, 25,\, 0.2\}$$

graphically proves that the range $[0.025;\, 10]$ of PTSDR did enclose the global minimum of the function (10). The best PDTM6080I-trained 4800-250-26-P by

$$\sigma_{\text{PD}}^{\langle\max\rangle} = r^{*}\cdot\sigma_{\text{TD}}^{\langle\max\rangle} = 0.225\cdot 0.2 = 0.045 \tag{22}$$

under set $\left\{C,\, F,\, Q_{\text{pass}}\right\} = \{2,\, 8,\, 25\}$ classifies TM6080I at CEP

$$p_{\text{EC}}(0.225) = \frac{1}{\sigma_{\text{TD}}^{\langle\max\rangle}} \int\limits_{\sigma_{\text{TD}}\in\left[0;\, \sigma_{\text{TD}}^{\langle\max\rangle}\right]} p_{\text{EC}}(0.225,\, \sigma_{\text{TD}},\, 1)\, d\sigma_{\text{TD}} = 5\int\limits_{0}^{0.2} p_{\text{EC}}(0.225,\, \sigma_{\text{TD}},\, 1)\, d\sigma_{\text{TD}} \approx$$

$$\approx \frac{1}{11}\sum_{t=0}^{10} p_{\text{EC}}(0.225,\, 0.02t,\, 1) < 0.28$$

and classifies PDTM6080I at CEP

$$p_{\text{EC}}(0.225) = \frac{1}{\sigma_{\text{TD}}^{\langle\max\rangle}} \int\limits_{\sigma_{\text{TD}}\in\left[0;\, \sigma_{\text{TD}}^{\langle\max\rangle}\right]} p_{\text{EC}}(0.225,\, \sigma_{\text{TD}},\, 2)\, d\sigma_{\text{TD}} = 5\int\limits_{0}^{0.2} p_{\text{EC}}(0.225,\, \sigma_{\text{TD}},\, 2)\, d\sigma_{\text{TD}} \approx$$

$$\approx \frac{1}{11}\sum_{t=0}^{10} p_{\text{EC}}(0.225,\, 0.02t,\, 2) < 0.48\,.$$

Hence, the problem (13), having been solved with $r^{*} = 0.225$ for a turned objects classification problem by 4800-250-26-P, opens a simple way for minimizing CEP over TM6080I and PDTM6080I. With the best PDTM6080I-trained 4800-250-26-P by (22) under set $\left\{C,\, F,\, Q_{\text{pass}}\right\} = \{2,\, 8,\, 75\}$ TM6080I are classified at CEP

$$p_{\text{EC}}(0.225) = \frac{1}{\sigma_{\text{TD}}^{\langle\max\rangle}} \int\limits_{\sigma_{\text{TD}}\in\left[0;\, \sigma_{\text{TD}}^{\langle\max\rangle}\right]} p_{\text{EC}}(0.225,\, \sigma_{\text{TD}},\, 1)\, d\sigma_{\text{TD}} = 5\int\limits_{0}^{0.2} p_{\text{EC}}(0.225,\, \sigma_{\text{TD}},\, 1)\, d\sigma_{\text{TD}} \approx$$

$$\approx \frac{1}{11}\sum_{t=0}^{10} p_{\text{EC}}(0.225,\, 0.02t,\, 1) < 0.19 \tag{23}$$

and PDTM6080I are classified at CEP

$$p_{\text{EC}}(0.225) = \frac{1}{\sigma_{\text{TD}}^{\langle\max\rangle}} \int\limits_{\sigma_{\text{TD}}\in\left[0;\, \sigma_{\text{TD}}^{\langle\max\rangle}\right]} p_{\text{EC}}(0.225,\, \sigma_{\text{TD}},\, 2)\, d\sigma_{\text{TD}} = 5\int\limits_{0}^{0.2} p_{\text{EC}}(0.225,\, \sigma_{\text{TD}},\, 2)\, d\sigma_{\text{TD}} \approx$$

$$\approx \frac{1}{11}\sum_{t=0}^{10} p_{\text{EC}}(0.225,\, 0.02t,\, 2) < 0.29\,, \tag{24}$$

and the averaged CEP by (11) is $p_{\text{EC}}(0.225) < 0.24$.

See in figure 16 how deep TD can be in an MI, that 4800-250-26-P by the optimized PTSDR still recognizes it. Up to brilliant, all MI in figure 16 have been classified correctly with PDTM6080I-trained 4800-250-26-P by with $r^{*} = 0.225$ under set $\left\{C,\, F,\, Q_{\text{pass}}\right\} = \{2,\, 8,\, 75\}$.

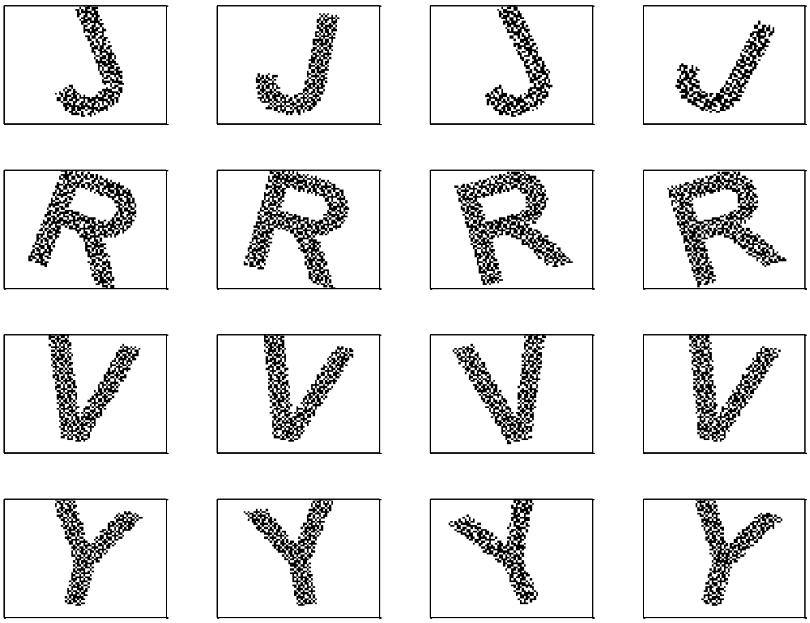

Figure 16. TM6080I of letters “J”, “R”, “V”, “Y” at $\sigma_{TD} = 0.2$ (views from within MATLAB), appeared recognizable for 2LP

Clearly that some day we shall absolutely recognize those letters in figure 1 also. Be it on perceptrons or neocognitrons, recognizability of suchlike objects must be at the level of performance of 4800-250-26-P, nearer to (23) or (24). It’s undisguised that if an object as 60×80 MI is rotated at any angle, then instead of GT with 26 classes NDR there would rather be GT with 104 or 208 classes. The last means that the initial object NDR splits into 4 or 8 independent NDR (figure 17). This doesn’t extend the spoken GT, but increases N_{classes}. Nonetheless 2LP will cope with the new classification problem, although it’ll require more neurons and IMR.

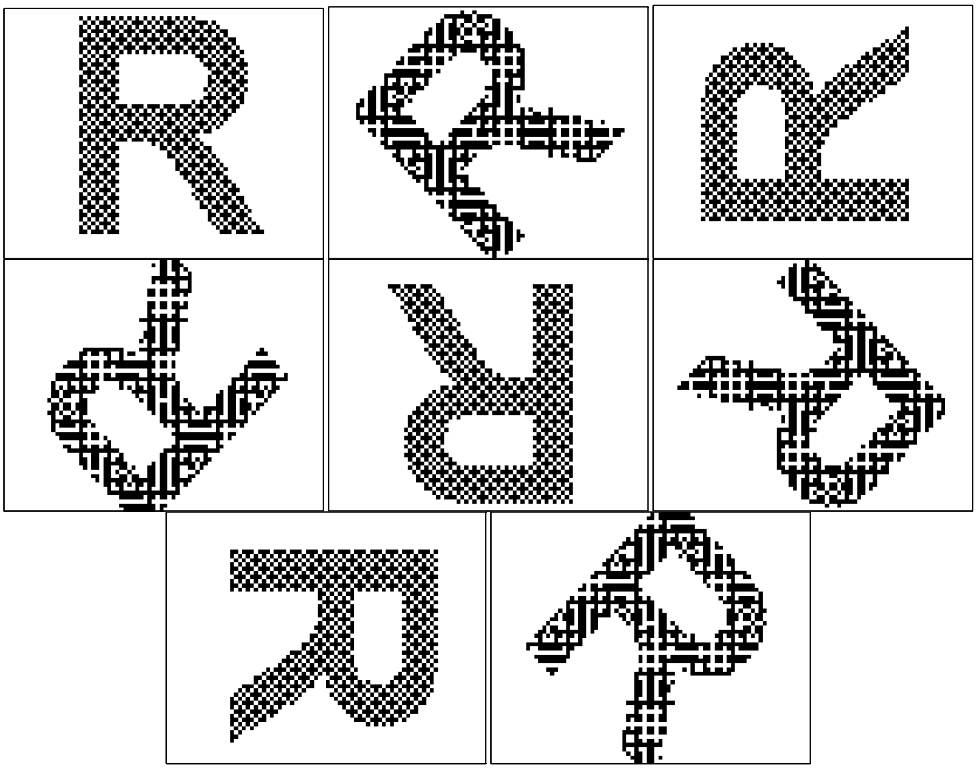

Figure 17. NDR of the letter «R» and TM6080I of the letter «R» at angles $\left\{\frac{\pi}{4}u\right\}_{u=1}^{7}$ from GT of 2^{4800} MI (views from within MATLAB), generating 8 different classes

INFERENCES

It's clear the problem (1) solution for training 4800-250-26-P on PDTM6080I optimally in turned objects classification problem can be immediately applied just to objects from GT of 2^{4800} MI with 26 classes NDR of enlarged English alphabet capital letters. For other GT, or other N_{classes}, or other object model the ratio r^* may be different, but goal and tasks within the third section of this article sustain the algorithm of how make the optimization of PTSDR in general. Note importantly, that there's no discussion about GT on the basis of the training set (4), as GT including matrices of the type (5) is infinite, it's a continuum (GT for training and GT for classification differ).

Optimization of other parameters of the training set (4) can minimize CEP even lower. Particularly, the raw integers C, F and Q_{pass} could be optimized for CEP minimization. At least, if CEP is practically tolerable at some $Q_{\text{pass}}^{**} < Q_{\text{pass}}^{*}$ by the optimal $Q_{\text{pass}} = Q_{\text{pass}}^{*}$ then reducing the passes number down to Q_{pass}^{**} optimizes IMR.

Another subsequent problem is to minimize CEP as the function of two variables — PTSDR and the size of SHL in 2LP. Its solution may shift the optimal PTSDR in relation to r^* by (13), and instead CEP will be lowered more, and the training process may be shortened along with IMR optimization owing to the optimal number of neurons in SHL. However, tabulating statistics for evaluating CEP as function of PTSDR and SHL size to plot surfaces is going to take far and away much more time in training and testing than this article has taken.

REFERENCES

1. Fukushima K. Artificial vision by multi-layered neural networks: Neocognitron and its advances / K. Fukushima // Neural Networks. — 2013. — Volume 37. — P. 103 — 119.
2. Plaza J. On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images / J. Plaza, A. Plaza, R. Perez, P. Martinez // Pattern Recognition. — 2009. — Volume 42, Issue 11. — P. 3032 — 3045.
3. Cireşan D. Multi-column deep neural network for traffic sign classification / D. Cireşan, U. Meier, J. Masci, J. Schmidhuber // Neural Networks. — 2012. — Volume 32. — P. 333 — 338.
4. Siniscalchi S. M. Exploiting deep neural networks for detection-based speech recognition / S. M. Siniscalchi, D. Yu, L. Deng, C.-H. Lee // Neurocomputing. — 2013. — Volume 106. — P. 148 — 157.
5. Kathirvalavakumar T. Neighborhood based modified backpropagation algorithm using adaptive learning parameters for training feedforward neural networks / T. Kathirvalavakumar, S. Jeyaseeli Subavathi // Neurocomputing. — 2009. — Volume 72, Issues 16 — 18. — P. 3915 — 3921.
6. Castillo P. A. Comparing evolutionary hybrid systems for design and optimization of multilayer perceptron structure along training parameters / P. A. Castillo, J. J. Merelo, M. G. Arenas, G. Romero // Information Sciences. — 2007. — Volume 177, Issue 14. — P. 2884 — 2905.
7. Kuzmanovski I. Counter-propagation neural networks in Matlab / I. Kuzmanovski, M. Novič // Chemometrics and Intelligent Laboratory Systems. — 2008. — Volume 90, Issue 1. — P. 84 — 91.
8. Ballabio D. A MATLAB toolbox for Self Organizing Maps and supervised neural network learning strategies / D. Ballabio, M. Vasighi // Chemometrics and Intelligent Laboratory Systems. — 2012. — Volume 118. — P. 24 — 32.

9. Nied A. 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. — Volume 70, Issue 16 — 18. — P. 2687 — 2691.

10. Yoo S. J. Indirect adaptive control of nonlinear dynamic systems using self recurrent wavelet neural networks via adaptive learning rates / S. J. Yoo, J. B. Park, Y. H. Choi // Information Sciences. — 2007. — Volume 177, Issue 15. — P. 3074 — 3098.

11. Romanuke V. V. Setting the hidden layer neuron number in feedforward neural network for an image recognition problem under Gaussian noise of distortion / V. V. Romanuke // Computer and Information Science. — 2013. — Volume 6, No. 2. — P. 38 — 54.

12. Романюк В. В. Зависимость производительности нейросети с прямой связью с одним скрытым слоем нейронов от гладкости её обучения на зашумленных копиях алфавита образов / В. В. Романюк // Вісник Хмельницького національного університету. Технічні науки. — 2013. — № 1. — С. 201 — 206.

REFERENCES

1. Fukushima, K. (2013), “Artificial vision by multi-layered neural networks: Neocognitron and its advances”, *Neural Networks*, vol. 37, pp. 103-119.

2. Plaza, J., Plaza, A., Perez, R. and Martinez, P. (2009), “On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images”, *Pattern Recognition*, vol. 42, issue 11, pp. 3032-3045.

3. Cireşan, D., Meier, U., Masci, J. and Schmidhuber, J. (2012), “Multi-column deep neural network for traffic sign classification”, *Neural Networks*, vol. 32, pp. 333-338.

4. Siniscalchi, S.M., Siniscalchi, S.M., Yu, D., Deng, L. and Lee, C.-H. (2013), “Exploiting deep neural networks for detection-based speech recognition”, *Neurocomputing*, vol. 106, pp. 148-157.

5. Kathirvalavakumar, T. and Jeyaseeli Subavathi S. (2009), “Neighborhood based modified backpropagation algorithm using adaptive learning parameters for training feedforward neural networks”, *Neurocomputing*, vol. 72, issues 16-18, pp. 3915-3921.

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

7. Kuzmanovski, I. and Novič, M. (2008), “Counter-propagation neural networks in Matlab”, *Chemometrics and Intelligent Laboratory Systems*, vol. 90, issue 1, pp. 84-91.

8. Ballabio, D. and Vasighi, M. (2012), “A MATLAB toolbox for Self Organizing Maps and supervised neural network learning strategies”, *Chemometrics and Intelligent Laboratory Systems*, vol. 118, pp. 24-32.

9. Nied, A., Seleme, S.I.Jr., Parma, G.G. and Menezes, B.R. (2007), “On-line neural training algorithm with sliding mode control and adaptive learning rate”, *Neurocomputing*, vol. 70, issue 16-18, pp. 2687-2691.

10. Yoo, S.J., Park, J.B. and Choi, Y.H. (2007), “Indirect adaptive control of nonlinear dynamic systems using self recurrent wavelet neural networks via adaptive learning rates”, *Information Sciences*, vol. 177, issue 15, pp. 3074-3098.

11. Romanuke, V.V. (2013), “Setting the hidden layer neuron number in feedforward neural network for an image recognition problem under Gaussian noise of distortion”, *Computer and Information Science*, vol. 6, no. 2, pp. 38-54.

12. Romanuke, V.V. (2013), “Dependence of performance of feed-forward neuronet with single hidden layer of neurons against its training smoothness on noised replicas of pattern alphabet”, *Bulletin of Khmelnitskiy National University. Technical Sciences*, no. 1, pp. 201-206.