ITHEA

International Journal

INFORMATION TECHNOLOGIES & KNOWLEDGE

2009 Volume 3 Number 1

International Journal
INFORMATION TECHNOLOGIES & KNOWLEDGE
Volume 3 / 2009, Number 1

Editor in chief: **Krassimir Markov** (Bulgaria)

International Editorial Board

Victor Gladun (Ukraine)

Abdelmgeid Amin Ali	**(Egypt)**	**Larissa Zaynutdinova**	**(Russia)**
Adil Timofeev	**(Russia)**	**Laura Ciocoiu**	**(Romania)**
Aleksey Voloshin	**(Ukraine)**	**Luis F. de Mingo**	**(Spain)**
Alexander Kuzemin	**(Ukraine)**	**Martin P. Mintchev**	**(Canada)**
Alexander Lounev	**(Russia)**	**Natalia Ivanova**	**(Russia)**
Alexander Palagin	**(Ukraine)**	**Nelly Maneva**	**(Bulgaria)**
Alfredo Milani	**(Italy)**	**Nikolay Lyutov**	**(Bulgaria)**
Avram Eskenazi	**(Bulgaria)**	**Orly Yadid-Pecht**	**(Israel)**
Axel Lehmann	**(Germany)**	**Peter Stanchev**	**(USA)**
Darina Dicheva	**(USA)**	**Radoslav Pavlov**	**(Bulgaria)**
Ekaterina Solovyova	**(Ukraine)**	**Rafael Yusupov**	**(Russia)**
Eugene Nickolov	**(Bulgaria)**	**Rumyana Kirkova**	**(Bulgaria)**
George Totkov	**(Bulgaria)**	**Stefan Dodunekov**	**(Bulgaria)**
Hasmik Sahakyan	**(Armenia)**	**Stoyan Poryazov**	**(Bulgaria)**
Ilia Mitov	**(Bulgaria)**	**Tatyana Gavrilova**	**(Russia)**
Irina Petrova	**(Russia)**	**Vadim Vagin**	**(Russia)**
Ivan Popchev	**(Bulgaria)**	**Vasil Sgurev**	**(Bulgaria)**
Jeanne Schreurs	**(Belgium)**	**Velina Slavova**	**(Bulgaria)**
Juan Castellanos	**(Spain)**	**Vitaliy Lozovskiy**	**(Ukraine)**
Julita Vassileva	**(Canada)**	**Vladimir Lovitskii**	**(UK)**
Karola Witschurke	**(Germany)**	**Vladimir Ryazanov**	**(Russia)**
Koen Vanhoof	**(Belgium)**	**Zhili Sun**	**(UK)**

IJ ITK is official publisher of the scientific papers of the members of the ITHEA International Scientific Society

IJ ITK rules for preparing the manuscripts are compulsory.
The **rules for the papers** for IJ ITK as well as the **subscription fees** are given on ***www.ithea.org***
The camera-ready copy of the paper should be received by e-mail: info@foibg.com.
Responsibility for papers published in IJ ITK belongs to authors.

General Sponsor of IJ ITK is the **Consortium FOI Bulgaria** (www.foibg.com).

International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Vol.3, Number 1, 2009

Edited by the **Institute of Information Theories and Applications FOI ITHEA®**, Bulgaria,
in collaboration with the **V.M.Glushkov Institute of Cybernetics of NAS, Ukraine**,
and the **Institute of Mathematics and Informatics, BAS, Bulgaria.**

Publisher: **ITHEA®**
Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org, e-mail: info@foibg.com
Printed in Bulgaria

Copyright © 2009 All rights reserved for the publisher and all authors.
® 2007-2009 "Information Technologies and Knowledge" is a trademark of Krassimir Markov

ISSN 1313-0455 (printed) **ISSN 1313-048X (online)** **ISSN 1313-0501 (CD/DVD)**

ADAPTIVE COMPARTMENTAL WAVELON WITH ROBUST LEARNING ALGORITHM

Yevgeniy Bodyanskiy, Oleksandr Pavlov, Olena Vynokurova

***Abstract**: In this paper a robust learning algorithm for adaptive compartmental wavelon based on R. Welsh criterion is proposed. Suggested learning algorithm under consideration allows the signals processing in presence of significant noise level and outliers. The robust learning algorithm efficiency is investigated and confirmed by the number of experiments including bio-medical applications.*

***Keywords**: computational intelligence, hybrid architecture, wavelet, adaptive compartmental wavelon, robust learning algorithm, outliers resistant.*

***ACM Classification Keywords**: I.2.6 Learning – Connectionism and neural nets.*

Introduction

Nowadays artificial neural networks (ANN) have gained the significant prevalence for solving the wide class of the information processing problems, uppermost for the identification, emulation, intelligent control, time series forecasting of arbitrary kind under significant noise level, and also the structural and parametric uncertainty.

The multilayer feedforward networks of three-layer perceptron type, where the elementary nodes are so-called P - neurons with monotonic activation functions are the most known and popular. The efficiency of the multilayer networks is explained by their universal approximation properties in combination with relative compact presentation of the simulated nonlinear system. It means, that they can be used successfully in the simulation (emulation) of non-linear systems tasks, which can be described by the equation

$$y(k) = F(x(k)) + \xi(k), \tag{1}$$

where $y(k)$ is the output system signal in k -th instant of discrete time $k = 0,1,2,\dots$, $x(k) \in X$ - $(n \times 1)$ is the vector of input signal, including both exogenous variables and previous values of the output signal, $F(\bullet)$ is the arbitrary function, generally in some unknown form, $\xi(k)$ is the unobserved disturbance with unknown characteristics. Usually it is assumed that function $F(\bullet)$ is defined either on the unit hypercube or on the orthotop

$$x_i(k) \in [x_i^{\min}, x_i^{\max}],\ i = 1,2,\dots,n,$$

where $x_i^{\min}$, $x_i^{\max}$ are the known low and upper limits of the i -th input influence variation.

The principal disadvantage of the multilayer networks is the low learning rate which is based on backpropagation algorithm which makes their application in the real time tasks impossible.

Alternative to the multilayer ANNs are the radial basis function networks, having one hidden layer consisting of, so-called, R -neurons. These networks learning is realized on the level of the output layer which is usually represented by the adaptive linear associator [Moody, Darken, 1989; Moody, Darken, 1988; Park, Sandberg, 1991; Leonard, 1992; Sunil, Yung, 1994; Poggio, Girosi, 1994]. Unlike P-neurons, R-neurons conventionally have bell-shaped activation function $f_j(x)$, where the argument is a distance (usually in Euclidean metric) between the current value of input signal $x(k)$ and the center c_j of the j-th neuron, i.e.

$$\varphi_j(x(k)) = \varphi_j\left(\sum_{i=1}^{n}(x_i(k)-c_{ji})^2\right) = \varphi_j\left(\left\|x(k)-c_j\right\|^2\right) \quad (2)$$

The principal advantage of RBFN is the high learning rate in the output layer, because the tuning parameters are linearly included to the network description. At the same time the problem of R-neurons centers allocation is remaining, and its unsuccessful solving leads to the «curse of dimensionality» problem. Using clustering techniques though allows reducing the size of the network, but excludes the possibility of on-line operation. Here it can be noted, that in [Bishop, 1995] the gradient recurrent procedure of the component-wise tuning parameters c_{ji} is described, but it is characterized by the low learning rate.

Along with neural networks for the arbitrary type signals processing, in the last years the wavelet theory is used sufficiently often [Chui, 1992; Daubechies, 1992], providing the compact local signal presentation both in the frequency and time domains. At the turn of the artificial neural network and wavelets theories the wavelet neural networks [Billings, 2005; Zhang, Benveniste, 1992; Zhang, 1997; Bodyanskiy, Vynokurova et al, 2006; Bodyanskiy, Vynokurova et al, 2007; Bodyanskiy et al, 2007; Bodyanskiy, Vynokurova et al, 2008] have evolved their efficiency for the analysis of non-stationary nonlinear signals and processes.

Elementary nodes of the wavelet neural networks are so-called radial wavelons [Reyneri, 1999], where the activation functions are the even wavelets with argument in form the Euclidian distance between $x(k)$ and wavelet translation vector c_j, where that every component of distance $\left|x_i(k)-c_{ji}\right|$ is weighted by the dilation parameter σ_{ji} such, that

$$\varphi_j(x(k)) = \varphi_j\left(\sum_{i=1}^{n}\left((x_i(k)-c_{ji})/\sigma_{ji}\right)^2\right) \quad (3)$$

where $\varphi_j(\bullet)$ is wavelet activation function. The receptive fields for such wavelons are hyperellipsoids with axes which are collinear to coordinate axes of the space X.

Taking into consideration the equivalence of radial basis ANN and fuzzy inference systems [Jang, Sun, 1993; Hunt et al, 1996], and also possibility of using even wavelet as a membership function [Mitaim, Kosko, 1996; Mitaim, Kosko 1997], within the bounds of the unification paradigm [Reyneri, 1999] we can talk about such hybrid system as adaptive compartmental wavelon having the radial-basis function network fast learning ability, fuzzy inferences systems interpretability and wavelet’s local properties.

It can be noted, that mostly tuning algorithms based on traditional squared learning criteria in the case of the processing data being contaminated by outliers with unknown distribution law, have shown themselves very sensitive to anomalous outliers. Thus the actual task is a synthesis of the robust learning algorithms, thet allow signal processing in presence of anomalous outliers.

This paper is devoted to synthesis of robust learning algorithm for adaptive compartmental wavelon, which has adjustable level of insensitivity to the different kind of outliers, rough errors, non-Gaussian disturbances, has high convergence rate and provides the advanced approximation properties in comparison with conventional computational intelligence systems. This hybrid structure can be used as the nodes of the neural network architectures.

1. Adaptive Compartmental Wavelon Structure

Let us consider the two-layers architecture shown on fig. 1 that coincides with the traditional radial-basis neural network. The input layer of the architecture is the receptor and in current time instant k the input signal in vector form $x(k) = (x_1(k), x_2(k), \ldots, x_n(k))^T$ is fed on it. Unlike radial basis function network the hidden layer consists of not by R-neurons, but by wavelons with wavelet activation function in the form

$$\varphi_j(x(k)) = \varphi_j\left(\left((x(k)-c_j)^T Q_j^{-1}(x(k)-c_j)\right), \alpha_j\right), \quad j = 1, 2, \ldots, h \tag{4}$$

in which instead of translation parameters σ_{ji} in (1) the dilation matrix Q_j is used, i.e. it is not Euclidian distance, but Itakura-Saito metric [Itakura, 1975].

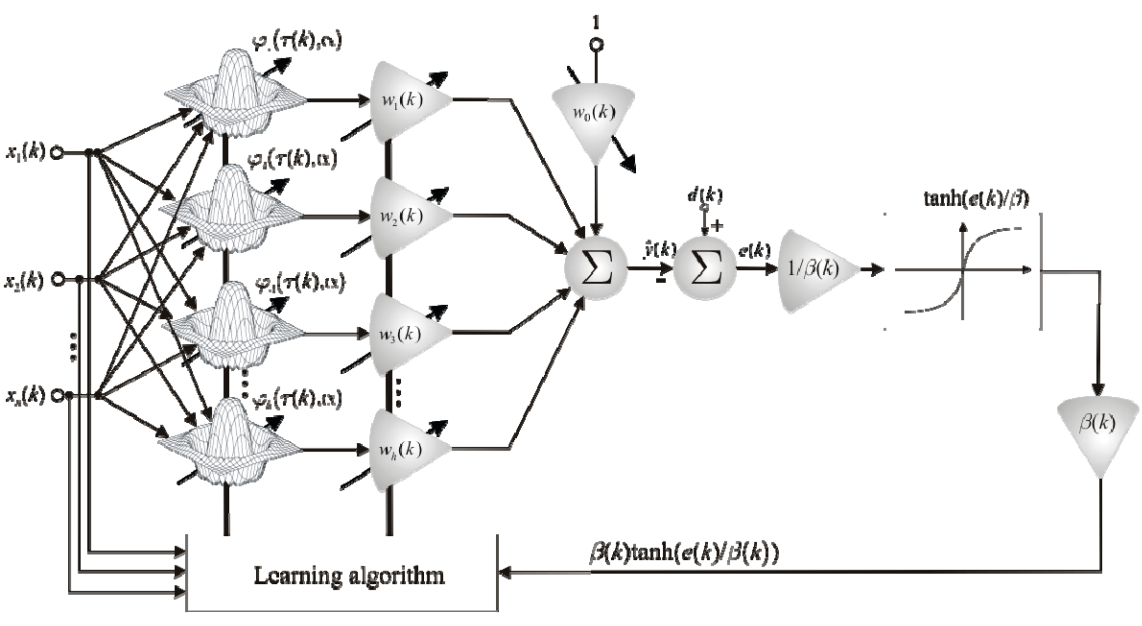


Fig. 1 – Adaptive compartmental wavelon with robust block

This results to the fact that receptive fields – wavelons hyperellipsoids (2) can have the arbitrary orientation relatively to the coordinate axes of space X, what extends the functional properties of adaptive compartmental wavelon.

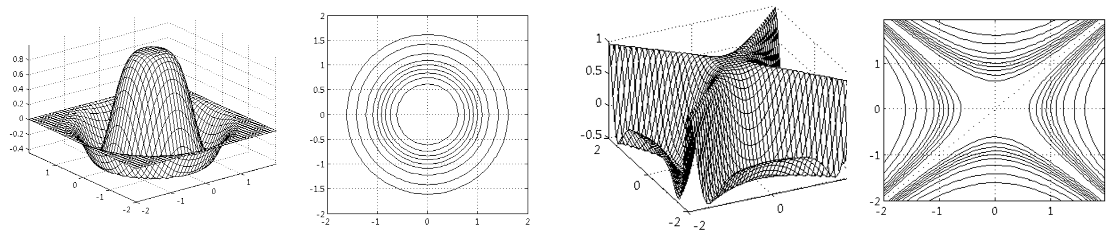
Based on the results of the author [Mitaim, Kosko, 1996; Mitaim, Kosko 1997], about that the wavelet-function can be used as a membership function in fuzzy systems, we can introduce the adaptive membership function based on wavelet Mexican Hat, having form

$$\varphi_j(\tau_j(k)) = (1-\alpha_j \tau_j^{\,2}(x(k)))\exp\left(-\frac{\tau_j^{\,2}(x(k))}{2}\right) \qquad (5)$$

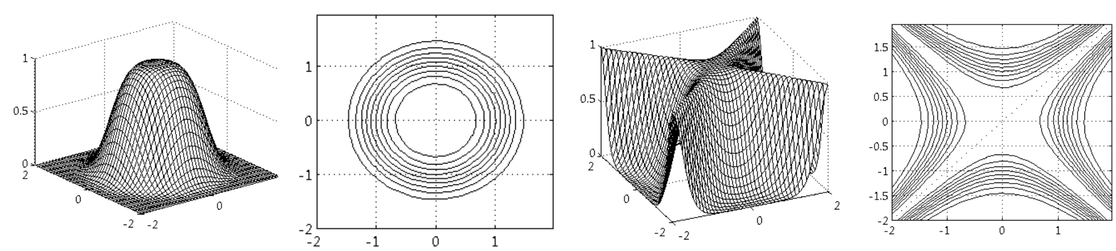
where $\tau_j(x(k)) = \left((x(k)-c_j(k))^T Q_j^{-1}(k)(x(k)-c_j(k))\right)$, α_j is turning parameter ($0 \le \alpha \le 1$).

Adaptive parameter α_j allows to tune the form of membership function in process of hybrid architecture learning, thus if $\alpha = 0$ then we get Gauss membership function, if $\alpha = 1$ then we get wavelet membership function Mexican Hat, and if $0 < \alpha < 1$ then we get hybrid membership function.

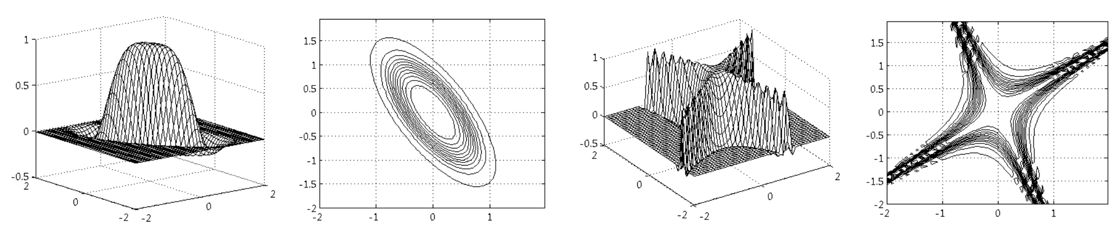
Fig. 2 shows the wavelon activation function (5) with arbitrary matrices Q_j and parameter α_j .



a) $Q^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \alpha = 1$ b) $Q^{-1} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \alpha = 1$



c) $Q^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \alpha = 0$ d) $Q^{-1} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \alpha = 0$



f) $Q^{-1} = \begin{pmatrix} 4 & 3 \\ 1 & 2 \end{pmatrix}, \alpha = 0.5$ g) $Q^{-1} = \begin{pmatrix} -4 & 3 \\ 1 & 2 \end{pmatrix}, \alpha = 0.5$

Fig. 2 – Wavelon activation function with arbitrary matrices Q_j and parameter α_j

And at last, the output layer is the common adaptive linear associator with tuning synaptic weights w_j

$$\hat{y}(k) = w_0(k) + \sum_{j=1}^{h} w_j(k)\varphi\Big(\big((x(k)-c_j)^T Q_j^{-1}(x(k)-c_j)\big), \alpha_j(k)\Big) = w(k)^T \varphi(\tau(k),\alpha), \qquad (6)$$

where $\varphi_0(\tau(k)) \equiv 1$, $w(k) = (w_0(k), w_1(k), w_2(k), \ldots, w_h(k))^T$, $\varphi(\tau(k),\alpha) = (1, \varphi_1(\tau_1(k),\alpha_1), \varphi_2(\tau_2(k),\alpha_2),$ $\ldots, \varphi_h(\tau_h(k),\alpha_h))^T$, $\tau(k) = (x(k)-c_j)^T Q_j^{-1}(x(k)-c_j)$.

Thus the tuning parameters of architecture to be determined in the learning process form the set of the $h+1$ synaptic weights w_j, h $(n\times 1)$-vectors c_j, h $(n\times n)$-matrices Q_j^{-1} and h $(n\times 1)$-vectors α_j. In total such network includes h $(1+2n+n^2)+1$ adjustable parameters.

2. The Robust Learning Algorithm for Adaptive Compartmental Wavelon

The experience shows that the identification methods based on the least square criterion are extremely sensitive to the deviation of real data distribution law from Gaussian distribution. In presence of various type outliers, an outrage errors, and non-Gaussian disturbances with “heavy tails” the methods based on the least squares criterion loose their efficiency.

In this case the methods of robust estimation and identification [Rey, 1978] which have obtained the wide spread for the learning of the artificial neural networks [Cichocki, Lobos, 1992; Cichocki, Unbehauen, 1993; Li, Chen, 2002] appear on the first role.

Let’s introduce into the consideration the learning error

$$e(k) = y(k) - \hat{y}(k) = y(k) - w^T(k)\varphi(k) \qquad (7)$$

and robust identification criterion by R. Welsh [Holland, Welsh, 1977; Welsh, 1977]

$$E(k) = f(k) = \beta^2 \ln\big(\cosh\big(e(k)/\beta\big)\big) \qquad (8)$$

where β is a positive parameter, that is chosen from empirical reasons and defining the size of zone of tolerance to outliers. It is necessary to note, that robust criterion (8) satisfies to all metric space axioms.

Fig. 3 shows the comparison of the robust optimization criterion with the least squares criterion.

Further we shall consider the learning algorithms synthesis. For the synaptic weights and the waveleon parameters (vectors c_j and matrices Q_j^{-1}) tuning we use gradient minimization of criterion (9), thus unlike the component-wise learning considered in [Bishop, 1995], we make some corrections in the vector-matrix form, that, firstly is easier from computing point of view, and secondly it allows to optimize learning process on the operation rate.

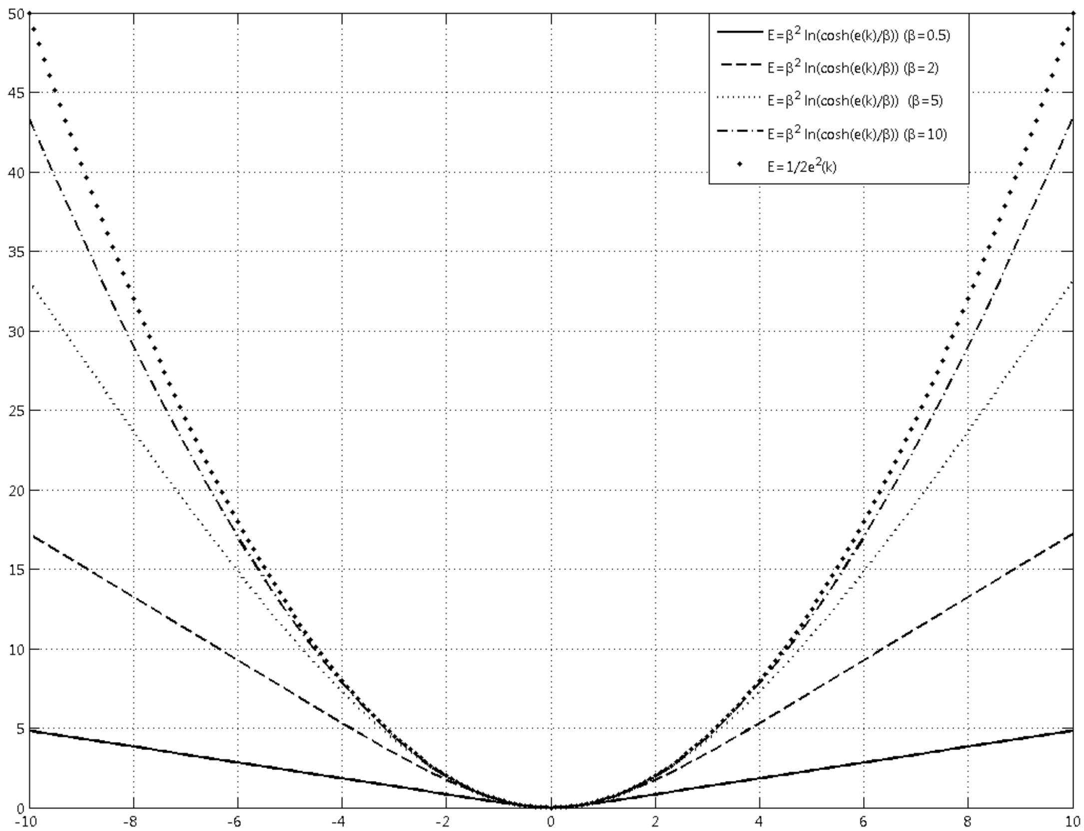


Fig. 3 – Comparison of the optimization criteria

In general case the learning algorithm can be written in form

$$\begin{cases} w(k+1)=w(k)-\eta_w\nabla_w E(k), \\ c_j(k+1)=c_j(k)-\eta_{c_j}\nabla_{c_j}E(k),\ \ j=1,2,\ldots,h, \\ Q_j^{-1}(k+1)=Q_j^{-1}(k)-\eta_{Q_j^{-1}}\left\{\partial E(k)\big/\partial Q_j^{-1}\right\},\ \ j=1,2,\ldots,h, \\ \alpha(k+1)=\alpha(k)-\eta_\alpha\nabla_\alpha E(k), \end{cases} \tag{9}$$

where $\nabla_w E$ is vector-gradient of the criterion (8) on w, $\nabla_{c_j}E$ is $(n\times 1)$ -vector-gradient criterion (9) on c_j; $\left\{\partial E(k)\big/\partial Q_j^{-1}\right\}$ is $(n\times n)$ -matrix, formed by partial derivatives $E(k)$ on components Q_j^{-1}; $\nabla_\alpha E$ is $(n\times 1)$ -vector-gradient criterion (9) on α; η_w, η_{c_j}, $\eta_{Q_j^{-1}}$, η_α are the learning rates.

For the wavelet (5) we can write

$$\begin{cases} \nabla_w E(k) = -\beta \tanh\left(e(k)/\beta\right)(1-\alpha_j \tau_j{}^2(x(k))) \exp\left(-\frac{\tau_j{}^2(x(k))}{2}\right) = -\tanh\left(e(k)/\beta\right) J_w(k), \\ \nabla_{c_j} E(k) = \beta \tanh\left(e(k)/\beta\right) w_j(k)\left(\alpha\tau_j^3(x(k)) - (2\alpha+1)\tau_j(x(k))\right)\cdot \\ \cdot\exp\left(-\frac{\tau_j^2(x(k))}{2}\right) Q_j^{-1}(k)(x(k)-c_j(k)) = \tanh\left(e(k)/\beta\right) J_{c_j}(k), \\ \left\{\partial E(k)/\partial Q_j^{-1}\right\} = -\beta \tanh\left(e(k)/\beta\right) w_j(k)\left(\alpha\tau_j^3(x(k)) - (2\alpha+1)\tau_j(x(k))\right)\cdot \\ \cdot\exp\left(-\frac{\tau_j^2(x(k))}{2}\right)(x(k)-c_j(k))(x(k)-c_j(k))^T = -\tanh\left(e(k)/\beta\right) J_{Q_j^{-1}}(k), \\ \nabla_\alpha E(k) = -\beta \tanh\left(e(k)/\beta\right) w(k)\tau^2(x(k)) \exp\left(-\frac{\tau^2(x(k))}{2}\right) = \tanh\left(e(k)/\beta\right) J_\alpha(k), \end{cases} \quad (10)$$

where $\tau_j(x(k)) = (x(k)-c_j(k))^T Q_j^{-1}(k)(x(k)-c_j(k))$.

Then the wavelons learning algorithm of the hidden layer subject to (10) is taking the form

$$\begin{cases} w(k+1) = w(k) + \eta_w \tanh\left(e(k)/\beta\right) J_w(k), \\ c_j(k+1) = c_j(k) - \eta_{c_j} \tanh\left(e(k)/\beta\right) J_{c_j}(k), \\ Q_j^{-1}(k+1) = Q_j^{-1}(k) + \eta_{Q_j^{-1}} \tanh\left(e(k)/\beta\right) J_{Q_j^{-1}}(k), \\ \alpha(k+1) = \alpha(k) + \eta_\alpha \tanh\left(e(k)/\beta\right) J_\alpha(k), \end{cases} \quad (11)$$

at that convergence rate to the optimal value w, c_j, Q_j^{-1} and α is completely defined by learning rate parameters η_w, η_{c_j}, $\eta_{Q_j^{-1}}$, and η_α.

The learning rate increasing can be achieved by using procedures more complex than gradient ones, such as Hartley or Marquardt procedures, that for the first relation (11) can be written in general form [Bodyanskiy, 1987; Bodyanskiy, Vynokurova et al, 2008]

$$w(k+1) = w(k) - \lambda_w (J_w(k) J_w^T(k) + \eta_w I)^{-1} J_w(k) \tanh\left(e(k)/\beta\right), \quad (12)$$

where I is the $(n \times n)$-identity matrix, λ_w is a positive dampening parameter, η_w is a momentum term parameter.

Using the inverse matrices lemma and after applying simple transformations we obtain the effective parameters learning algorithm in the form

$$
\begin{cases}
w(k+1) = w(k) + \lambda_w \left(\tanh\left(e(k)/\beta\right) J_w(k)\right) \Big/ \left(\eta_w + \left\|J_w(k)\right\|^2\right), \\
c_j(k+1) = c_j(k) - \lambda_c \left(\tanh\left(e(k)/\beta\right) J_{c_j}(k)\right) \Big/ \left(\eta_c + \left\|J_{c_j}(k)\right\|^2\right), \\
Q_j^{-1}(k+1) = Q_j^{-1}(k) + \lambda_{Q_j^{-1}} \left(\tanh\left(e(k)/\beta\right) J_{Q_j^{-1}}(k)\right) \Big/ \left(\eta_{Q_j^{-1}} + Tr(J^T_{Q_j^{-1}}(k) J_{Q_j^{-1}}(k))\right), \\
\alpha(k+1) = \alpha(k) + \lambda_\alpha \left(\tanh\left(e(k)/\beta\right) J_\alpha(k)\right) \Big/ \left(\eta_\alpha + \left\|J_\alpha(k)\right\|^2\right).
\end{cases}
\tag{13}
$$

In order to add more smoothing properties, using approach proposed in [Bodyanskiy et al, 2001; Bodyanskiy, Vynokurova et al, 2008], we can introduce the modified learning procedure:

$$
\begin{cases}
w(k+1) = w(k) + \lambda_w \dfrac{\tanh\left(e(k)/\beta\right) J_w(k)}{\eta_w(k)}, & \eta_w(k+1) = \gamma\eta_w(k) + \left\|J_w(k+1)\right\|^2, \\
c_j(k+1) = c_j(k) - \lambda_{c_j} \dfrac{\tanh\left(e(k)/\beta\right) J_{c_j}(k)}{\eta_{c_j}(k)}, & \eta_{c_j}(k+1) = \alpha_c\eta_{c_j}(k) + \left\|J_{c_j}(k+1)\right\|^2, \\
Q_j^{-1}(k+1) = Q_j^{-1}(k) + \lambda_{Q_j^{-1}} \dfrac{\tanh\left(e(k)/\beta\right) J_{Q_j^{-1}}(k)}{\eta_{Q_j^{-1}}(k)}, & \eta_{Q_j^{-1}}(k+1) = \gamma\eta_{Q_j^{-1}}(k) + Tr\left(J^T_{Q_j^{-1}}(k+1) J_{Q_j^{-1}}(k+1)\right) \\
\alpha(k+1) = \alpha(k) + \lambda_\alpha \dfrac{\tanh\left(e(k)/\beta\right) J_\alpha(k)}{\eta_\alpha(k)}, & \eta_\alpha(k+1) = \gamma\eta_\alpha(k) + \left\|J_\alpha(k+1)\right\|^2,
\end{cases}
\tag{14}
$$

(here $0 \le \gamma \le 1$ are the parameters of weighting out-dated information), being nonlinear hybrid of the Kaczmarz-Widrow-Hoff and Goodwin-Ramadge-Caines algorithms and including both following and filtering properties.

3. Results of the Experimental Research

In the first experiment the developed robust learning algorithm was tested out on the basis of a signal with intensive outliers. The signal had been obtained using Narendra's nonlinear dynamical system (it is a standard benchmark, widely used to evaluate and compare the performance of neural and neuro-fuzzy systems for nonlinear system modeling and time series forecasting) whose output signal is artificially contaminated by random noise generated according to the Cauchy distribution with the inverse transform method described by equation in form

$$F_X^{-1}(x) = x_0 + \gamma tg\left[\pi\left(x - 0.5\right)\right] \quad (15)$$

where x_0 is the location parameter, γ is the scale parameter $(\gamma > 0)$, x is the support area ($x \in (-\infty, +\infty)$).

The nonlinear dynamical system is generated by equation in form [Narendra, Parthasarathy, 1990]

$$y(k+1) = 0.3y(k) + 0.6y(k-1) + f(u(k)), \quad (16)$$

where $f(u(k)) = 0.6\sin(\pi u(k)) + 0.3\sin(3\pi u(k)) + 0.1\sin(5\pi u(k))$ and $u(k) = \sin(2\pi k/250)$, $0 < k < 20000$ is discrete time. The values $x(t-2), x(t-1), x(t)$ were used to emulate $x(t+1)$. In the on-line mode of learning, adaptive compartmental wavelon was trained with procedure (13) for 20000 iterations. The parameters of the learning algorithm were $\beta_w = 0.5, \beta_c = 0.5, \beta_Q = 0.5, \beta_\alpha = 0.5,$ $\gamma_w = \gamma_c = \gamma_Q = \gamma_\alpha = 0.99,$ $\lambda_w = \lambda_{c_j} = \lambda_{Q_j^{-1}} = \lambda_\alpha = 0.99$. Initial values were $\eta_w(0) = \eta_{c_j}(0) = \eta_{Q_j^{-1}}(0) = \eta_\alpha(0) = 10000$. After 20000 iterations the training was stopped, and the next 5000 points were used as the testing data set. Testing data set included $f(u(k)) = 0.6\sin(\pi u(k)) + 0.3\sin(3\pi u(k)) + 0.1\sin(5\pi u(k))$ and $u(k) = \sin(2\pi k/250)$ for $20001 < k < 22000$ and $f(u(k)) = u^3(k) + 0.3u^2(k) - 0.4u(k)$ and $u(k) = \sin(2\pi k/250) + \sin(2\pi k/25)$ for $22001 < k < 25000$.

Fig. 3 a shows the results of the noised nonstationary signal emulation (real values (dashed line) and emulated values (solid line)). Fig. 3 b shows segment of the learning process; as it can be seen the number of outliers with large amplitude, present in the beginning of the sample, didn't have a significant influence on the learning algorithm.

The comparison of emulation results based on robust learning algorithm with results of emulation based on gradient algorithm and the algorithm based on recurrent least squares method where the structure network and the number of tuning parameters were identical was carried out.

Under adaptive compartmental wavelon learning using the gradient algorithm the first outlier in the beginning of the sample, had a noticeable influence on the learning algorithm. Under adaptive compartmental wavelon learning using the recurrent least-squares method the first occurred outlier leads to the covariance matrix so-called "parameters blow-up" what results in inability to emulate signals noised by anomalous outliers. Thus it is obvious that the proposed robust learning algorithm allows signal processing under high level outliers noise conditions.

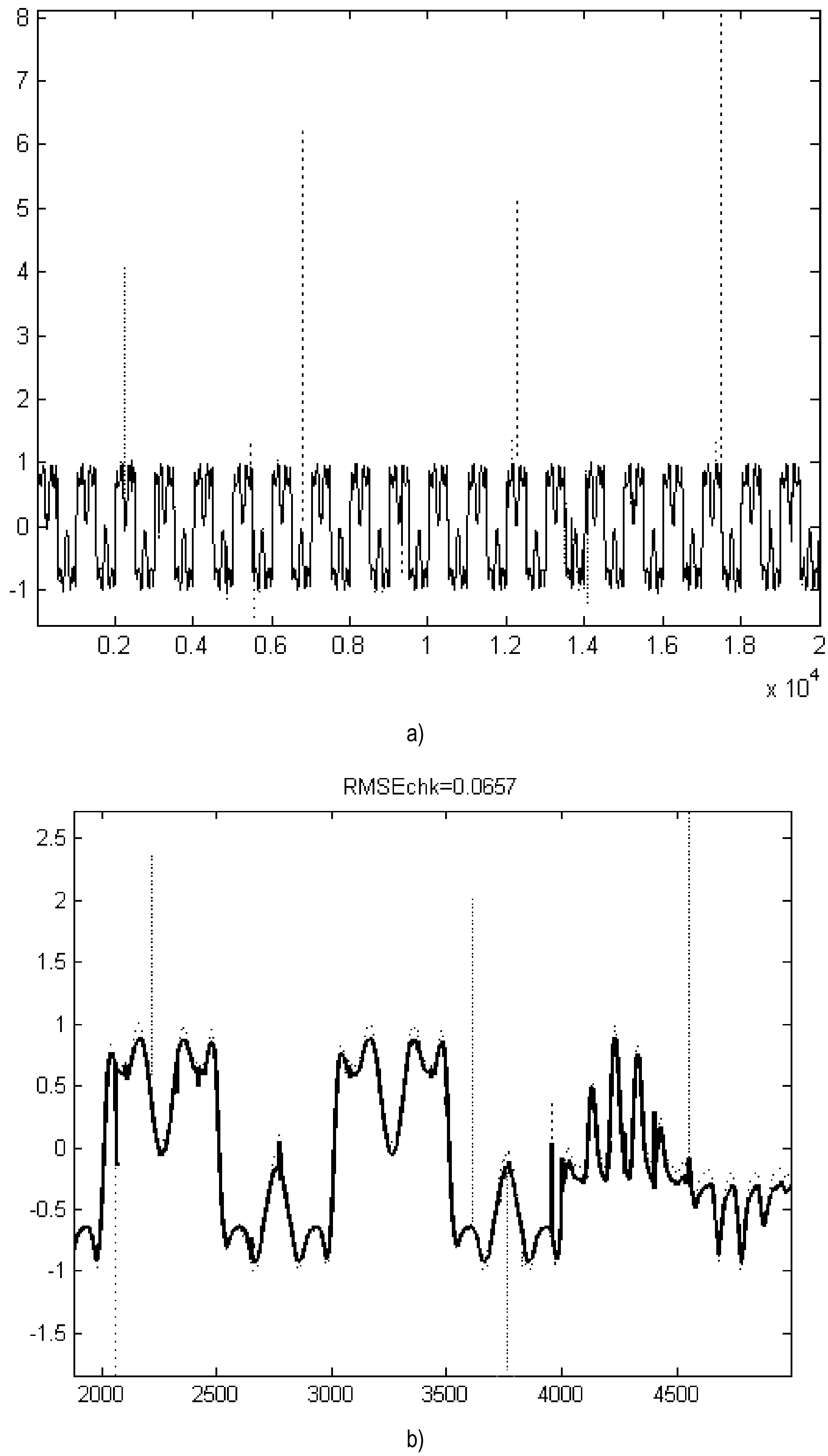


Fig. 3 – Results of noised nonstationary signal emulation based on robust learning algorithm

In Table 1 the comparison results are shown.

Table 1: The results of noisy signal emulation

Neural Network / Learning algorithm	RMSE
Adaptive compartmental wavelon / Proposed robust learning algorithm (13) with parameter α	0.0657
Adaptive compartmental wavelon / Proposed robust learning algorithm (13) without parameter α	0.0998
Adaptive compartmental wavelon / The gradient learning algorithm	1.1436
Adaptive compartmental wavelon / RLSM	∞

The second experiment has been made on the data set, presented by Government Institution "Institute of General and Urgent Surgery (Academy of Medical Sciences of Ukraine)". It has been carried out studying of the homeostasis indexes dynamic of the patient with the stomach acute injury [Cook et al, 1996; Vynokurova, Pavlov et al, 2007] based on outliers resistant radial-basis-fuzzy-wavelet-neural network. The indexes of oxygen cascade, system hemodynamics, daily pH- measurement, and hypoxia marker and endotoxemia were analyzed. Result of processing studied clinico-laboratory data set was the degree defining of enteral deficiency, that it has allowed to lead the adequate stomach-protect diagnosis and therapy.

Conclusion

In the paper computationally simple and effective all adaptive compartmental wavelon parameters robust learning algorithm is proposed. The robust learning algorithm has following and smoothing properties and allows on-line processing of nonstationary signals under a number of outliers and “heavy tails” disturbances. Addition of wavelons receptor fields, including their transformations (dilation, translation, rotation and form parameters of activation function) allows to improve the network approximation properties, that is confirmed by the experiments research results.

Bibliography

[Bishop, 1995] C. M. Bishop Neural Networks for Pattern Recognition. Oxford: Clarendon Press, 1995.

[Billings, 2005] S. A. Billings, H.-L. Wei A new class of wavelet networks for nonlinear system identification. IEEE Trans. on Neural Networks, 16, 2005, P. 862-874.

[Bodyanskiy, 1987] Ye. Bodyanskiy Adaptive algorithms for nonlinear control plant identification. ASU i pribory avtomatiki, Kharkiv: Vyscha shk., 81, 1987, P. 43-46. (In Russian)

[Bodyanskiy et al, 2001] Ye. Bodyanskiy, V. Kolodyazhniy, A. Stephan An adaptive learning algorithm for a neuro-fuzzy network. Computational Intelligence. Theory and Applications, Ed. by B. Reusch, Berlin-Heidelberg-New York: Springer, 2001, P. 68-75.

[Bodyanskiy, Vynokurova et al, 2006] Ye. Bodyanskiy, N. Lamonova, O. Vynokurova Recurrent learning algorithm for double-wavelet neuron. Proc. XII -th Int. Conf. "Knowledge - Dialogue - Solution", Varna, 2006, P.77-84.

[Bodyanskiy, Vynokurova et al, 2007] Ye. Bodyanskiy, N. Lamonova, O. Vynokurova Double-wavelet neuron based on analytical activation functions. Int. J. Information Theory and Applications., 14, 2007, P. 281-288.

[Bodyanskiy et al, 2007] Ye. Bodyanskiy, I. Pliss, O. Vynokurova A learning algorithm for forecasting adaptive wavelet-neuro-fuzzy network Proc. XIII -th Int. Conf. "Information Research & Applications", Varna, 2007, P.211-218.

[Bodyanskiy, Vynokurova et al, 2008] Ye. Bodyanskiy, O. Vynokurova, V.V. Boyko, S.V. Sushkov, O.O. Pavlov Clinical data analysis in the medical researches based on computational intelligence methods, Kharkiv, Eksklusiv, 2008. (In Russian)

[Chui, 1992] C. K. Chui An Introduction to Wavelets. New York: Academic Press, 1992.

[Cichocki, Lobos, 1992] A. Cichocki, T. Lobos Adaptive analogue network for real time estimation of basic waveforms of voltages and currents IEEE Proc, 139, Part C , 1992, P. 343-350.

[Cichocki, Unbehauen, 1993] A. Cichocki, R. Unbehauen Neural Networks for Optimization and Signal Processing. Stuttgart: Teubner, 1993.

[Cook et al, 1996] D.J. Cook, B.K. Reeve, G.H. Guyatt et al. Stress ulcer prophylaxis in critically ill patients. Resolving discordant meta-analyses. JAMA, 1996, 257, P. 308-314.

[Daubechies, 1992] I. Daubechies Ten Lectures on Wavelets. Philadelphia, PA: SIAM.,1992.

[Holland, Welsh, 1977] P.W. Holland, R.E. Welsh Robust regression using iteratively reweighted least squares. Commun. Statist., Theory and Methods., 46, 1977, P. 813 – 828.

[Hunt et al, 1996] K. J. Hunt, R. Haas, R. M. Smith Extending the functional equivalence of radial basis function networks and fuzzy inference systems. IEEE Trans. on Neural Networks, 7, 1996, P. 776-781.

[Itakura, 1975] F. Itakura Maximum prediction residual principle applied to speech recognition. IEEE Trans. on Acoustics, Speech, and Signal Processing, 23, 1975, P. 67-72.

[Jang, Sun, 1993] J. S. R. Jang, C.T. Sun Functional equivalence between radial basis function networks and fuzzy inference systems. IEEE Trans. on Neural Networks, 4, 1993, P. 156-159.

[Leonard, 1992] J. A. Leonard, M. A. Kramer, L. H. Ungar Using radial basis functions to approximate a function and its error bounds. IEEE Trans. on Neural Networks, 3, 1992, P. 614-627.

[Li, Chen, 2002] S.-T. Li, S.-C. Chen Function approximation using robust wavelet neural networks. Proc. of the 14th IEEE Int. Conf. on Tools with Artificial Intelligence, 2002, P. 483- 488.

[Mitaim, Kosko, 1996] S. Mitaim, B. Kosko What is the best shape for a fuzzy set in function approximation? Proc. 5th IEEE Int. Conf on Fuzzy Systems “Fuzz-96”, V. 2, 1996, P. 1237-1213.

[Mitaim, Kosko 1997] S. Mitaim, B. Kosko Adaptive joint fuzzy sets for function approximation. Proc. Int. Conf. on Neural Networks “ICNN-97”, 1997, P. 537-542.

[Moody, Darken, 1989] J. Moody, C. Darken Fast learning in networks of locally-tuned processing units. Neural Computation, 1, 1989, P. 281-294.

[Moody, Darken, 1988] J. Moody, C. Darken Learning with localized receptive fields. In: Proc. of the 1988 Connectionist Models Summer School, Eds. D. Touretzky, G. Hinton, T. Sejnowski, 1988, San Mateo: Morgan-Kaufmann, P. 133-143.

[Narendra, Parthasarathy, 1990] K.S. Narendra, K. Parthasarathy Identification and control of dynamic systems using neural networks. IEEE Trans. on Neural Networks, 1990, 1, P. 4-26.

[Park, Sandberg, 1991] J. Park, I. W. Sandberg Universal approximation using radial-basis-function networks. Neural Computation, 3, 1991, P. 246-257.

[Poggio, Girosi, 1994] T. Poggio, F. Girosi A Theory of Networks for Approximation and Learning. A. I. Memo № 1140, C.B.I.P. Paper № 31., Massachussetts Institute of Technology, 1994.

[Rey, 1978] W.J.J. Rey Robust Statistical Methods. Lecture Notes in Mathematics, Berlin-Heidelberg-New York: Springer-Verlag, V. 690, 1978.

[Reyneri, 1999] L. M. Reyneri Unification of neural and wavelet networks and fuzzy systems. IEEE Trans. on Neural Networks, 1999, 10, P. 801-814.

[Sunil, Yung, 1994] E. V. T. Sunil, C. Sh. Yung Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems. IEEE Trans. on Neural Networks, 5, 1994, P. 594-603.

[Vynokurova, Pavlov et al, 2007] V.V. Boyko, S.V. Sushkov, O.O. Pavlov, O.A. Vynokurova The anesthesia and intensive therapy strategy in the acute gastro ileum hemorrhage, Kharkiv, Eksklusiv, 2007. (In Russian)

[Welsh, 1977] R.E. Welsh Nonlinear statistical data analysis. Proc. Comp. Sci. and Statist., Tenth Ann Symp. Interface. Held at Nat'l Bur. Stds. Gaithersburg, MD, 1977, P. 77-86.

[Zhang, Benveniste, 1992] Q. H. Zhang, A. Benveniste Wavelet networks. IEEE Trans. on Neural Networks, 3, 1992, P. 889-898.

[Zhang, 1997] Q. H. Zhang Using wavelet network in nonparametric estimation. IEEE Trans. on Neural Networks, 8, 1997, P. 227-236.

Authors' Information

***Bodyanskiy Yevgeniy** - Doctor of Technical Sciences, Professor of Artificial Intelligence Department and Scientific Head of the Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, Lenina av. 14, Kharkiv, Ukraine 61166, Tel +380577021890, e-mail: bodya@kture.kharkov.ua*

Major Fields of Scientific Research: Hybrid Systems of Computational Intelligence

***Pavlov Oleksandr** - Candidate of Medical Sciences, Senior Researcher, Doctor-Anesthesiologist, Government Institution "Institute of General and Urgent Surgery (Academy of Medical Sciences of Ukraine)", Balakireva av.,1, Kharkiv, Ukraine, 61018, Tel. +380577021890, e-mail: pavlov73@list.ru*

Major Fields of Scientific Research: Anesthesiology, Computational Intelligence in Medical Problems

***Vynokurova Olena** – Candidate of Technical Sciences, Senior Researcher of the Control Systems Research Laboratory, Associate Professor of Information Technologies Security Department, Kharkiv National University of Radio Electronics, Lenina av. 14, Kharkiv, Ukraine, 61166, Tel +380577021890, e-mail: vinokurova@kture.kharkov.ua*

Major Fields of Scientific Research: Hybrid Wavelet-Fuzzy-Neuro-Techniques, Computational Intelligence.