HYBRID CASCADE NEURAL NETWORK BASED ON WAVELET-NEURON

Yevgeniy Bodyanskiy, Oleksandra Kharchenko, Olena Vynokurova

Abstract: *In the paper new hybrid cascade wavelet-neural network and its learning algorithm in batch and on-line mode are proposed. Such architecture can be used for solving prediction and emulation non-stationary non-linear time series under current and a-priori uncertenity. The computational experiments confirm the effectiveness of developed approach.*

Keywords: *cascade neural network, wavelet-neuron, learning algorithm, prediction, emulation.*

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

Introduction

As part of an evolutionary approach to the synthesis of neural networks architecture it can be provided such direction as the cascade neural networks [Avedjan, 1999; Bodyanskiy, 2004a; Bodyanskiy, 2006; Bodyanskiy, 2007; Bodyanskiy, 2008a; Fahlman, 1990; Schalkoff, 1997]. The cascade-correlation neural network, which proposed S. Fahlman and C. Lebiere [Fahlman, 1990], is the most typical and effective representative of such neural networks. The main feature of this network type is the ability to add new nodes during learning process.

Thus the cascade neural networks are the flexible and effective approach for solving of the wide type tasks related to data mining under full and partial, a-priori and current uncertainty. The cascade neural networks are fully adaptive evolutionary architectures, because they tune not only synaptic weight but and cascade number during learning process. Fitting cascade number allows to select the architecture with complexity which is most suitable for the solving problem without the experts intervention in the object domain.

The main advantages of cascade-correlation networks are following ones:

- such networks do not demand a-priori defining both architecture network and neuron number in the cascades;
- neurons are added to the network as necessary, creating no hidden layers, but cascades, each of them uses the input of network and previous cascade outputs as the own input signals;

- the learning process is not associated with concept of back-propagation, that allows significantly to reduce the learning process time;
- the computational burden on the learning process is reduced by "freezing" of synaptic weights that formed previous cascades.

The main disadvantage of such networks is impossibility of their learning process in on-line mode [Bodyanskiy, 2004a], due to the type of used artificial neurons – elementary Rosenblatt perceptrons [Rosenblatt, 1964; Rosenblatt, 1966]. As is known in such neurons the sigmoidal or hyperbolic tangent functions are used as activation functions and as a result of the output signal of each neuron depends nonlinearly from the synaptic weights. So learning process should be performed using the delta-rule and its modifications, which are gradient optimization algorithms. Obviously, it is difficult to talk about optimizing of the learning rate and learning process in on-line mode in this case.

In the connection with that it seems appropriate to synthesize the hybrid cascade architecture, where the cascades use neurons, in which the output signal linearly depends of the synaptic weights that permits to optimize learning speed and reduce the size of the training set.

Wavelet-neuron and its learning algorithm

Let us consider the structure of the wavelet-neuron [Bodyanskiy, 2003; Bodyanskiy, 2004b; Bodyanskiy, 2005] shown in fig. 1. Evidently, wavelet-neuron is sufficiently close to structure of conventional n-input formal neuron, but instead of tuning synaptic weights contains wavelet synapses WS_i, $i = 1, 2, \ldots, n$, whose adjustable parameters are not only weights w_{ji}, but also the center and width of wavelet activation function $\varphi_{ji}(x_i(k))$.

When vector signal $x(k) = (x_1(k), x_2(k), \ldots, x_n(k))^T$ (where $k = 0, 1, 2, \ldots$ is discrete current time) is fed to the input of the wavelet-neuron, the output is determined by both the tunable weights $w_{ji}(k)$ and wavelet functions:

$$y(k) = \sum_{i=1}^{n} f_i(x(k)) = \sum_{i=1}^{n}\sum_{j=1}^{h} w_{ji}(k)\varphi_{ji}(x_i(k)). \quad (1)$$

Notice that wavelet-neuron architecture coincides with the neo-fuzzy neuron of T. Yamakawa [Miki, 1999; Uchino, 1997; Yamakawa, 1992], but differs in that, instead of triangular membership functions are used even wavelets in nonlinear synapses. However, as shown by B. Kosko [Mitaim, 1997], the use of even wavelets does not contradict the ideas of fuzzy inference, and the specific values of wavelet functions can be given a sense of membership levels.

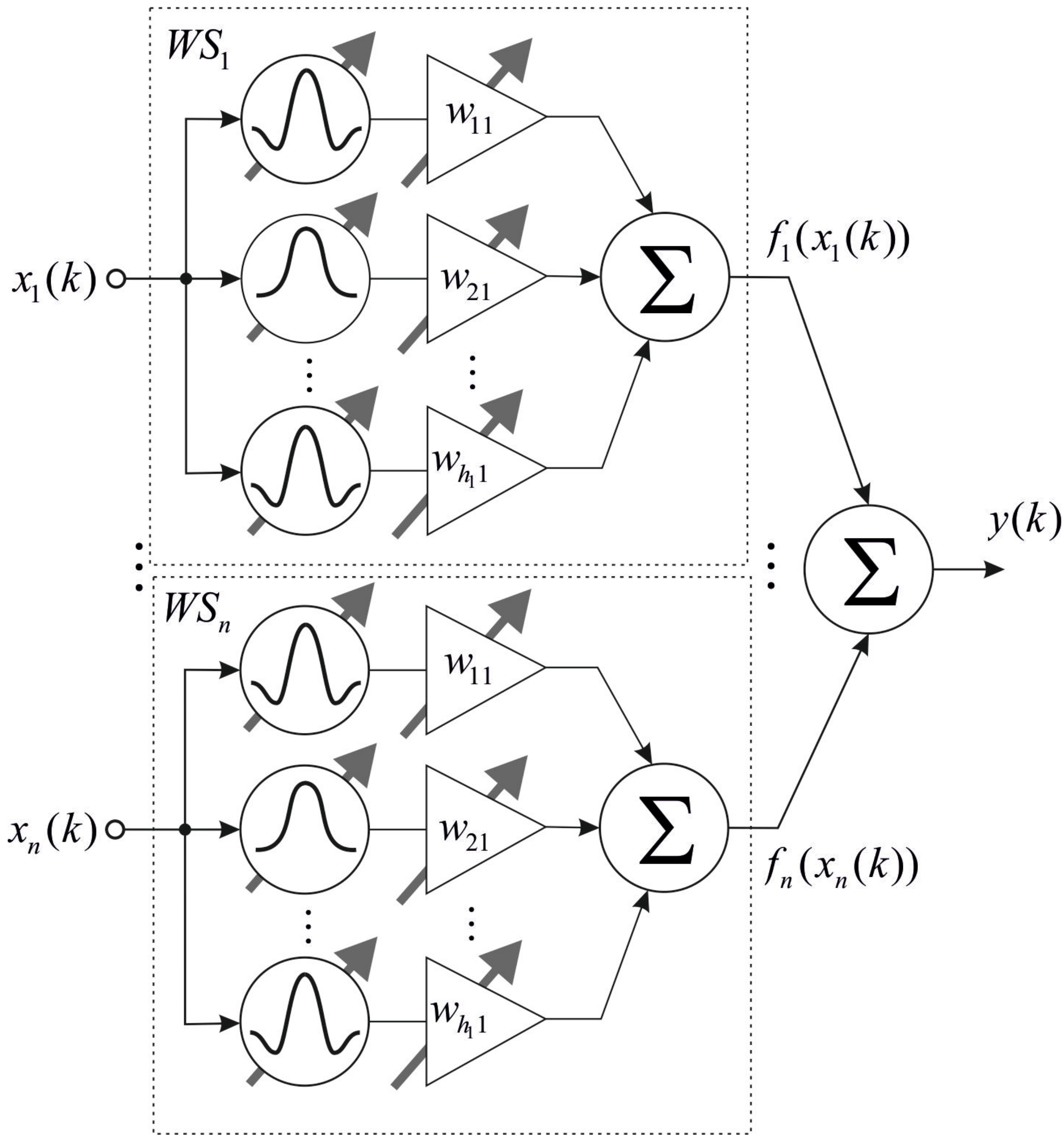

Figure 1. Wavelet-neuron architecture

The different wavelet-function type can be used as the activation function of wavelet-neuron. The most suitable function is proposed by us - adaptive wavelet activation-membership function [Bodyanskiy, 2008b] which has the form

$$\varphi_{ji}(x_i(k)) = (1-\alpha_{ji}(k)\tau_{ji}^2)\exp\left(-\tau_{ji}^2(k)/2\right), \tag{2}$$

where $\tau_{ji}(k) = (x_i(k) - c_{ji}(k))\sigma_{ji}^{-1}(k)$, $c_{ji}(k), \sigma_{ji}(k)$ are parameters, which define the center, width and $\alpha_{ji}(k)$ is parameter of function shape.

In [Bodyanskiy, 2003] the enough simple and effective learning algorithm of wavelet-neuron is proposed, which has the form

$$w_{ji}(k+1) = w_{ji}(k) + \eta^{w}(k)e(k)(1-\alpha_{ji}(k)\tau_{ji}(k))\exp\left(-\tau_{ji}^{2}(k)/2\right), \qquad (3)$$

where scalar coefficient $\eta^{w}(k)$ defines the step in the tuning parameters space.

The rate of convergence of the learning algorithm (3) can be increased via using of the second-order procedures, such as the Levenberg-Marquardt algorithm [Shepherd, 1997] which are widely used to train neural networks.

Introducing $(h_i \times 1)$-vectors of variables $\varphi_i(x_i(k)) = (\varphi_{1i}(x_i(k)),\ldots,\varphi_{h_i i}(x_i(k)))^T$, $w_i(k) = (w_{1i}(k),\ldots,w_{h_i i}(k))^T$ and $\tau_i(k) = (\tau_{1i}(k),\ \ldots,\tau_{h_i i}(k))^T$, we can obtain the gradient-based learning algorithm of the *i*-th wavelet synapse WS_i [Bodyanskiy, 2005; Otto, 2003]:

$$\begin{cases} w_i(k+1) = w_i(k) + \dfrac{e(k)\varphi_i(x_i(k))}{\eta^{w} + \|\varphi_i(x_i(k))\|^2} = w_i(k) + \dfrac{e(k)\varphi_i(x_i(k))}{\eta_j^{w}(k)}, \\ \eta_i^{w}(k+1) = \beta\eta_i^{w}(k) + \|\varphi_i(x_i(k))\|^2, \end{cases} \qquad (4)$$

which has both tracking and filtering properties, where β is a forgetting factor ($0 \le \beta \le 1$).

Hybrid cascade wavelet neural network

Replacing neurons in the nodes of the cascade network by the elements, whose output signals are linearly dependent of the synaptic weight we can escape disadvantages and even get some useful properties due to the choice of artificial neuron type.

Replacing Rosenblatt perceptrons in nodes of cascade correlation neural networks which proposed Fahlman and Lebier on the wavelet-neurons we can to introduce hybrid wavelet cascade architecture shown on the fig. 2.

Such cascade wavelet-neural network is realized mapping of the following form

– wavelet-neuron of first cascade

$$\hat{y}^{[1]} = \sum_{i=1}^{n}\sum_{j=1}^{h}\varphi_{ji}^{[1]}\mu_{ji}(x_i), \qquad (5)$$

– wavelet-neuron of second cascade

$$\hat{y}^{[2]} = \sum_{i=1}^{n}\sum_{j=1}^{h} w_{ji}^{[2]}\varphi_{ji}(x_i) + \sum_{j=1}^{h} w_{j,n+1}^{[2]}\varphi_{j,n+1}(\hat{y}^{[1]}), \qquad (6)$$

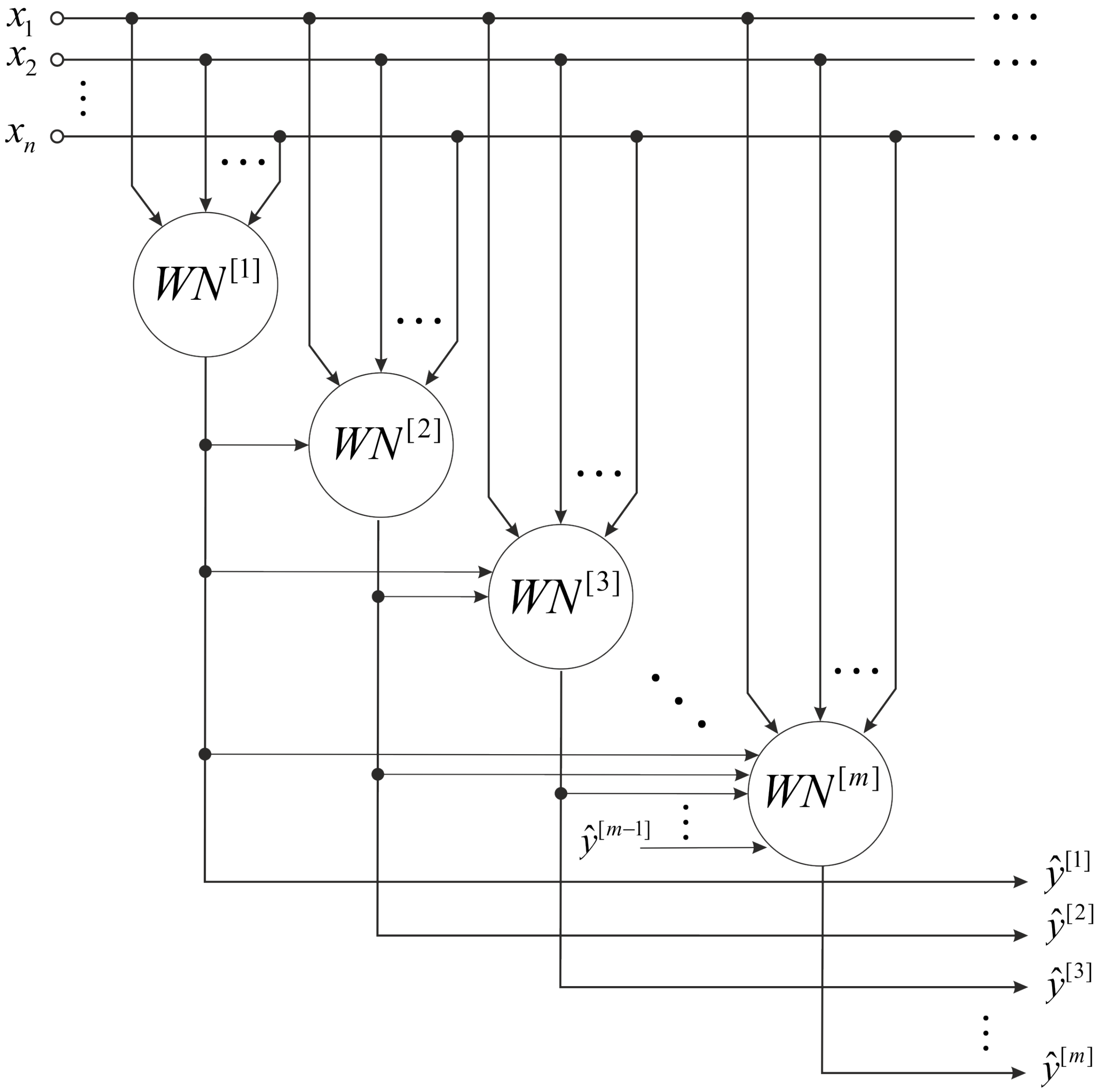

Figure 2. The cascade wavelet neural network

– wavelet-neuron of third cascade

$$\hat{y}^{[3]} = \sum_{i=1}^{n}\sum_{j=1}^{h} w_{ji}^{[3]}\varphi_{ji}(x_i) + \sum_{j=1}^{h} w_{j,n+1}^{[3]}\varphi_{j,n+1}(\hat{y}^{[1]}) + \sum_{j=1}^{h} w_{j,n+2}^{[3]}\varphi_{j,n+2}(\hat{y}^{[2]}), \tag{7}$$

– wavelet-neuron of m-th cascade

$$\hat{y}^{[m]} = \sum_{i=1}^{n}\sum_{j=1}^{h} w_{ji}^{[m]}\varphi_{ji}(x_i) + \sum_{l=n+1}^{n+m-1}\sum_{j=1}^{h} w_{jl}^{[m]}\varphi_{jl}(\hat{y}^{[l-n]}). \tag{8}$$

Leaning algorithms for hybrid cascade wavelet–neural network

The cascade wavelet neural network learning is performed in the batch mode using full training set $\{x(1), y(1); x(2), y(2);...; x(k), y(k);...; x(N), y(N)\}$.

At the beginning a set of wavelet-functions values (2) $\varphi^{[1]}(1), \varphi^{[1]}(2),..., \varphi^{[1]}(N)$ is calculated for each training sample. For the learning process of cascade wavelet-neural network the parameters of wavelet activation-membership function $c_{ji}(k), \sigma_{ji}(k)$ and $\alpha_{ji}(k)$ are defined by clustering procedure or can be defined based on uniform grid.

Then using direct minimization of the learning criterion

$$E_N^{[1]} = \frac{1}{2}\sum_{k=1}^{N} e_1(k)^2 = \frac{1}{2}\sum_{k=1}^{N}(y(k) - \hat{y}_1(k))^2, \tag{9}$$

the vector of synaptic weights can be computed as

$$w^{[1]}(N) = \left(\sum_{k=1}^{N}\varphi^{[1]}(k)\varphi^{[1]T}(k)\right)^{+}\sum_{k=1}^{N}\varphi^{[1]}(k)y(k) = P^{[1]}(N)\sum_{k=1}^{N}\varphi^{[1]}(k)y(k), \tag{10}$$

where $(\bullet)^{+}$ is symbol of Moore-Penrose pseudoinverse.

If dimension of this vector is sufficiently large it is suitable to use procedure based on recursive least squares method with sequential training samples processing:

$$\begin{cases} w^{[1]}(k+1) = w^{[1]}(k) + \dfrac{P^{[1]}(k)(y(k+1) - w^{[1]T}(k)\varphi^{[1]}(k+1))}{1 + \varphi^{[1]T}(k+1)P^{[1]}(k)\varphi^{[1]}(k+1)}\varphi^{[1]}(k+1), \\ P^{[1]}(k+1) = P^{[1]}(k) - \dfrac{P^{[1]}(k)\varphi^{[1]}(k+1)\varphi^{[1]T}(k+1)P^{[1]}(k)}{1 + \varphi^{[1]T}(k+1)P^{[1]}(k)\varphi^{[1]}(k+1)},\ P^{[1]}(0) = \gamma I, \end{cases} \tag{11}$$

where γ is sufficiently large positive number which defined empirically, I is unity matrix (of appropriate dimensionality).

It is necessary to notice that using procedures (10), (11) for adjusting weight coefficients essentially reduces learning time in comparison with gradient algorithms underlying delta-rule.

After first cascade learning completion, synaptic weights of the neuron WN_1 become 'frozen' and second cascade of network consisting from a single neuron WN_2 is generated. It has one additional input for the output signal of the first cascade. Then procedures (10), (11) again are applied for adjusting of weight coefficients vector $w^{[2]}$, with dimensionality $(h+1)(n+1)\times 1$.

The neural network growing process (increasing cascades number) continues until we obtain required precision of the solved problem's solution, and for the adjusting weight coefficients of the last *m*-th cascade following expression are used:

$$w^{[m]}(N)=\left(\sum_{k=1}^{N}\varphi^{[m]}(k)\varphi^{[m]T}(k)\right)^{+}\sum_{k=1}^{N}\varphi^{[m]}(k)y(k)=P^{[m]}(N)\sum_{k=1}^{N}\varphi^{[m]}(k)y(k) \tag{12}$$

in batch mode,

$$\begin{cases} w^{[m]}(k+1)=w^{[m]}(k)+\dfrac{P^{[m]}(k)(y(k+1)-w^{[m]T}(k)\varphi^{[m]}(k+1))}{1+\varphi^{[m]T}(k+1)P^{[m]}(k)\varphi^{[m]}(k+1)}\varphi^{[m]}(k+1), \\ P^{[m]}(k+1)=P^{[m]}(k)-\dfrac{P^{[m]}(k)\varphi^{[m]}(k+1)\varphi^{[m]T}(k+1)P^{[m]}(k)}{1+\varphi^{[m]T}(k+1)P^{[m]}(k)\varphi^{[m]}(k+1)} \end{cases} \tag{13}$$

or

$$\begin{cases} w^{[m]}(k+1)=w^{[m]}(k)+\dfrac{e(k)\varphi^{[m]}(x(k))}{\eta^{[m]}(k)}, \\ \eta^{[m]}(k+1)=\beta\eta^{[m]}(k)+\left\|\varphi^{[m]}(x(k))\right\|^{2},\ 0\le\beta\le 1 \end{cases} \tag{14}$$

in on-line mode of information processing.

The main disadvantage of conventional cascade-correlation network is their ability of the batch mode learning usage, when all training set should be given a-priori. Cascade wavelet neural network can be trained in on-line mode, because of algorithm (13), (14) possess maximal possible squared rate of convergence. In this case at the first step architecture consisting of *m* cascades is generated. Each cascade trains using proper algorithm. Since outputs of the previous wavelet-neurons become additional inputs for the *m*-th cascade, algorithm realizes recurrent method of the prediction error, well known in the theory of adaptive identification. Changing cascades number during learning process also can be easily performed.

Conclusion

In the paper hybrid cascade wavelet neural network is proposed. It differs from its prototype, cascade-correlation learning architecture, in increased speed of operation, numerical stability and real-time processing possibility. Theoretical justification and experiment results confirm the efficiency of developed approach.

Acknowledgement

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Bibliography

[Avedjan, 1999] E.D. Avedjan, G.V. Barkan, I.K. Levin Cascade Neural Networks. Avtomatika i Telemekhanika, 3, 1999. P.38-55. (In Russian)

[Bodyanskiy, 2000] Ye. Bodyanskiy, O. Mykhalev, I. Pliss Adaptive fault detection in control plants using artificial neural networks. Dnipropetrovsk: Systemni tekhnologiy, 2000. (In Ukrainian)

[Bodyanskiy, 2003] Ye. Bodyanskiy, O. Vynokurova Adaptive wavelet-neuro-predictor. Problemy bioniki, 58, 2003, P. 10-17. (In Russian)

[Bodyanskiy, 2004a] Ye. Bodyanskiy, O. Rudenko Artificial neural networks: architectures, learning and application. Kharkiv:TELETEKH, 2004. (In Russian)

[Bodyanskiy, 2004b] Ye. Bodyanskiy, V. Kolodyazhniy, I. Pliss, O. Vynokurova Learning wavelet neuron based on the RASP-function // Radioelectronics. Informatics. Control, 1 (11), 2004, P. 118-122.

[Bodyanskiy, 2005] Ye. Bodyanskiy, N. Lamonova, I. Pliss, O. Vynokurova An adaptive learning algorithm for a wavelet neural network. Blackwell Synergy: Expert Systems, 22(5), 2005, P. 235-240.

[Bodyanskiy, 2006] Ye. Bodyanskiy, I. Pliss, O. Slipchenko Growing neural networks based on orthogonal activation functions. Proc. XII-th Int. Conf. ITA KDS. Varna, 2006. P84-89.

[Bodyanskiy, 2007] Ye. Bodyanskiy, I. Pliss, O. Slipchenko Growing neural network using nonconventional activation functions. Int. J. Information Theories & Applications, 14, 2007. P.275-281.

[Bodyanskiy, 2008a] Ye. Bodyanskiy, A. Dolotov, I. Pliss, Ye. Viktorov The cascade orthogonal neural network. Advanced Research in Artificial Intelligence. Bulgaria, Sofia: Institute of Informational Theories and Applications FOI ITHEA., 2, 2008, P. 13-20.

[Bodyanskiy, 2008b] Ye. Bodyanskiy, O. Vynokurova, E. Yegorova Radial-basis-fuzzy-wavelet-neural network with adaptive activation-membership function. Int. J. on Artificial Intelligence and Machine Learning, 8, 2008, P. 9-15.

[Fahlman, 1990] S.E. Fahlman, C. Lebiere The cascade-correlation learning architecture. Advances in Neural Information Processing Systems. Ed. D.S. Touretzky. San Mateo, CA: Morgan Kaufman, 1990. P.524-532.

[Miki, 1999] T. Miki, T. Yamakawa Analog implementation of neo-fuzzy neuron and its on-board learning. Computational Intelligence and Applications. Ed. N.E. Mastorakis. Piraeus: WSES Press, 1999.P.144-149.

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

[Otto, 2003] P. Otto, Ye. Bodyanskiy, V. Kolodyazhniy A new learning algorithm for a forecasting neuro-fuzzy network. Integrated Computer-Aided Engineering. Amsterdam: IOS Press, 10 (4), 2003, P. 399-409.

[Rosenblatt, 1964] F. Rosenblatt The perceptron - a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory, 1957.

[Rosenblatt, 1966] F. Rosenblatt The perceptron: A probabilistic model for Information storage and organization in the brain, Cornel aeronautical laboratory. Psychological Review, 65 (6), 1958, P. 386-408.

[Schalkoff, 1997] R.J. Schalkoff Artificial Neural Networks. N.Y.: The McGraw-Hill Comp., Inc., 1997.

[Shepherd, 1997] A. J. Shepherd Second-Order Methods for Neural Networks. London: Springer–Verlag, 1997.

[Uchino, 1997] E. Uchino E., T. Yamakawa Soft computing based signal prediction, restoration and filtering. Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks and Genetic Algorithms. Ed. Da Ruan. Boston: Kluwer Academic Publisher, 1997. P.331-349.

[Yamakawa, 1992] T. Yamakawa, E. Uchino, T. Miki, H. Kusanagi A neo fuzzy neuron and its applications to system identification and prediction of the system behavior. Proc. 2-nd Int.Conf. on Fuzzy Logic and Neural Networks "IIZUKA-92". Iizuka, Japan, 1992. P.477-483.

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

Kharchenko Oleksandra – *Student, Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, Lenina av. 14, Kharkiv, Ukraine 61166, Tel +380577021890*

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

Vynokurova Olena – *Candidate of Technical Sciences, Leading 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.