EVOLVING CASCADE NEURAL NETWORK BASED ON MULTIDIMESNIONAL EPANECHNIKOV'S KERNELS AND ITS LEARNING ALGORITHM

Yevgeniy Bodyanskiy, Paul Grimm, Nataliya Teslenko

Abstract: *At present time neural networks based on Group Method of Data Handling (GMDH-NN), nodes of which are two-input N-Adalines, is well-known. Each of N-Adalines contains the set of adjustable synaptic weights that are estimated using standard least squares method and provides quadratic approximation of restoring nonlinear mapping. On the other hand, for needed approximation quality ensuring this NN can require considerable number of hidden layers. Approximating properties of GMDH-NN can be improved by uniting the approaches based on Group Method of Data Handling and Radial-Basis-Functions Networks that have only one hidden layer, formed by, so-called, R-neurons. Such networks learning reduces, as a rule, to the tuning of synaptic weights of output layer that are formed by adaptive linear associators. In contrast to neurons of multilayer structures with polynomial or sigmoidal activation functions R-neurons have bell-shaped activation functions. In this paper as activation functions multidimensional Epanechnikov's kernels are used. The advantage of activation function is that its derivatives are linear according all the parameters that allows to adjust sufficiently simply not only synaptic weights but also centers with receptive fields. Proposed network combines Group Method of Data Handling, Radial-Basis-Functions Networks and cascade networks and isn't inclined to the "curse of dimensionality", is able to real time mode information processing by adapting its parameters and structure to problem conditions. The multidimensional Epanechnikov's kernels were used as activation functions, that allowed to introduce numerically simple learning algorithms, which are characterized by high speed.*

Keywords: *evolving neural network, cascade networks, radial-basis neural network, Group Method of Data Handling, multidimensional Epanechnikov's kernels.*

ACM Classification Keywords: *F.1 Computation by abstract devices – Self-modifying machines (e.g., neural networks), I.2.6 Learning – Connectionism and neural nets, G.1.2 Approximation – Nonlinear approximation.*

Introduction

At present artificial neural networks have gotten a wide spread for extensive class of pattern recognition, identification, emulation, intelligent control, time series prediction etc. problems due to universal approximating properties and abilities to learn. As far as when a number of practical tasks solving the volume of learning sample is limited, then to the foreground learning rate factor comes. At the same time not all the neural networks are able to overcome arising problems and, first of all, so-called "overfitting". As one of the most high-performance networks that are learned based on optimization procedures of second order with high convergence rate is Radial Basis Functions Neural Network (RBFN). Output signal of this network linearly depends on adjusting synaptic weights. At the same time these networks are inclined to so-called "curse of dimensionality", when the number of radial-basis neurons of hidden layer (R-neurons) exponentially grows while input signals state grows.

It is possible to overcome this problem by dividing the initial task in one or another way to a number of subtasks of low dimensionality and grouping obtained solutions to get required result. From computational point of view the most convenient in this case is Group Method of Data Handling (GMDH) [Ivakhnenko, 1969; Ivakhnenko, 1975; Ivakhnenko, Stepashko, 1975; Ivakhnenko, Madala, 1994] that demonstrated its efficiency when solving a great number of practical tasks.

In [Pham, Liu, 1995] multi-layered GMDH-neural network was considered. It has two-inputs N-adalines as a nodes and output of each node is quadratic function of input signals. At the same time the synaptic weights of each neuron are defined in patch mode using standard least squares method. It can be needed some more quantity of hidden layers to provide necessary approximation quality. That is why on-line learning becomes impossible.

Hybrid architecture of artificial neural network based on ideas of GMDH and consequent forming of cascade neural networks [Avedyan, Barkan, Levin, 1999]. Nodes of this network are compartmental R-neurons with activation functions like multidimensional Epanechnikov's kernels [Epanechnikov, 1969; Friedman, Hastie, Tibshirani, 2003; Bodyanskiy, Chaplanov, Kolodyazhniy, Otto, 2002], that have large degree of freedom, and thus improved approximating properties in comparison with conventional Gaussians.

Compartmental R-neuron with Multidimensional Epanechnikov Kernels and Its Learning Algorithm

Let introduce the structure of compartmental R-neuron presented on fig.1 and concurring with simplified architecture of conventional Radial Basis Functions Neural Network with two inputs x_i and x_j, $i, j = 1, 2, \ldots, n$, where n – dimensionality of input space.

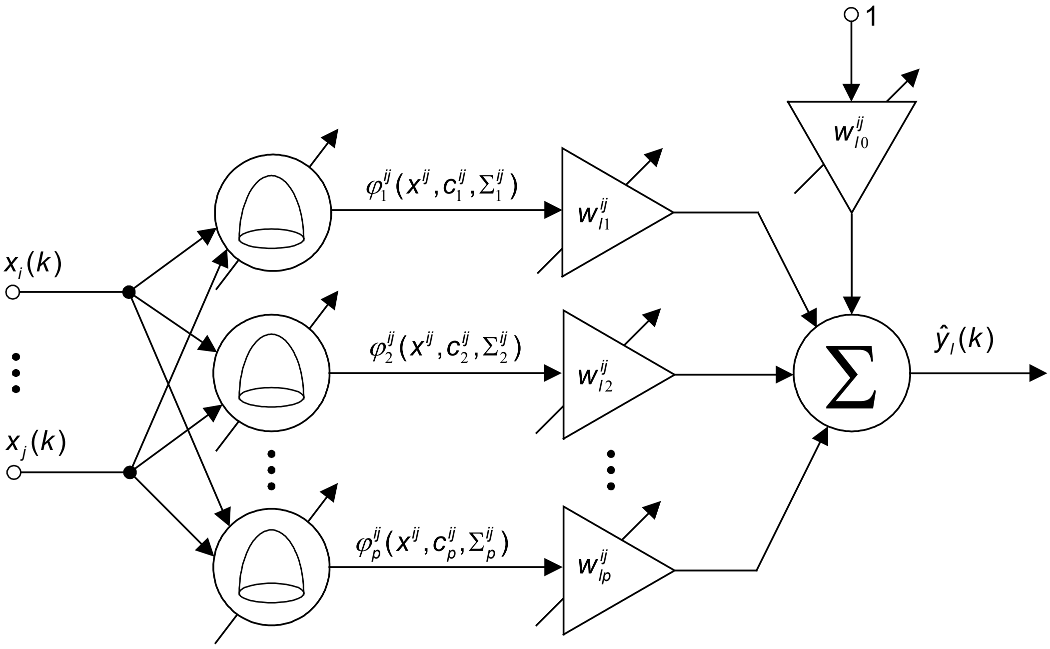


Fig. 1 – Compartmental R-neuron

Compartmental R-neuron contains p activation functions (conventionally in RBFN multidimensional Gaussians or other bell-shaped functions are used) $\varphi_h^{ij}(x^{ij}, c_h^{ij}, \Sigma_h^{ij})$, p+1 synaptic weights that are united to vector $w_l^{ij} = (w_{l0}^{ij}, w_{l1}^{ij}, \ldots, w_{lp}^{ij})$, p two-dimensional vector of centers $c_h^{ij} = (c_h^i, c_h^j)^{\mathrm{T}}$, p (2×2) – matrices of receptive fields of activation functions Σ_h^{ij},two-dimensional inputs vector $x^{ij}=(x_i,x_j)^{\mathrm{T}}$, one output $\hat{y}_l$; $l=1,2,\ldots,p$; $k=1,2,\ldots,N$ – number of observation in processing sample or index of current discrete time.

Multidimensional Epanechnikov's kernels are used as activation functions $\varphi_h^{ij}(x^{ij},c_h^{ij},\Sigma_h^{ij})$

$$\varphi_h^{ij}(x^{ij},c_h^{ij},\Sigma_h^{ij})=1-\left\|x^{ij}-c_h^{ij}\right\|^2_{\left(\Sigma_h^{ij}\right)^{-1}}, \tag{1}$$

that have bell-shaped by positive definite matrix of receptive field Σ_h^{ij}. The advantage of activation function (1) in comparison with conventional ones is in linearity of its derivatives with respect to all the parameters that allows to adjust not only synaptic weights but also centers with receptive fields sufficiently easy.

At the same time, transformation that is realized by compartmental R-neural has a form

$$\hat{y}_l=w_{l0}^{ij}+\sum_{h=1}^{p}w_{lh}^{ij}\varphi_h^{ij}(x^{ij},c_h^{ij},\Sigma_h^{ij})=w_{l0}^{ij}+\sum_{h=1}^{p}w_{lh}^{ij}\left(1-\left\|x^{ij}-c_h^{ij}\right\|^2_{\left(\Sigma_h^{ij}\right)^{-1}}\right).$$

Usually RBFN learning comes to synaptic weights w_{lh}^{ij} adjusting, but centers and characteristics of receptive fields are defined priory. At the same time for two-dimensional case it is sufficiently easy to locate the centers at the regular lattice nodes and define receptive field as circles. Learning process itself consists of synaptic weights vector w_l^{ij} estimating by learning sample containing N observations $x^{ij}(k)=(x_i(k),x_j(k))^{\mathrm{T}}$, $y(k)$, $k=1,2,\ldots,N$, where $y(k)$ - external learning signal.

By introducing to the consideration $(p+1)\times 1$-vector of activation functions $\varphi^{ij}(k)=(1,\varphi_1^{ij}(x^{ij}(k),c_1^{ij},\Sigma_1^{ij}),\ldots,\varphi_p^{ij}(x^{ij}(k),c_p^{ij},\Sigma_p^{ij}))^{\mathrm{T}}$ and learning criterion

$$E_l^N=\sum_{k=1}^{N}(y(k)-\hat{y}_l(k))^2=\sum_{k=1}^{N}e_l^2(k)=\sum_{k=1}^{N}(y(k)-(w_l^{ij})^{\mathrm{T}}\varphi^{ij}(k))^2, \tag{2}$$

using standard least squares method it is easy to obtain required solution in the form

$$w_l^{ij}=\left(\sum_{k=1}^{N}\varphi^{ij}(k)(\varphi^{ij}(k))^{\mathrm{T}}\right)^{+}\sum_{k=1}^{N}\varphi^{ij}(k)y(k), \tag{3}$$

where $(\bullet)^{+}$ – symbol of inversion by Moore-Penrose.

If the data are fed to the processing consequently in on-line mode, then instead of (3) can be used its recurrent variant in the form

$$\begin{cases} w_l^{ij}(k)=w_l^{ij}(k-1)+\dfrac{P_{ij}(k-1)(y(k)-(w_l^{ij}(k-1))^{\mathrm{T}}\varphi^{ij}(k))}{1+(\varphi^{ij}(k))^{\mathrm{T}}P_{ij}(k-1)\varphi^{ij}(k)}\varphi^{ij}(k), \\ P_{ij}(k)=P_{ij}(k-1)-\dfrac{P_{ij}(k-1)\varphi^{ij}(k)(\varphi^{ij}(k))^{\mathrm{T}}P_{ij}(k-1)}{1+(\varphi^{ij}(k))^{\mathrm{T}}P_{ij}(k-1)\varphi^{ij}(k)}, \qquad P_{ij}(0)=\gamma I, \qquad \gamma>>0. \end{cases} \tag{4}$$

Algorithms (3) and (4) are effective only in the cases, when required solution is stationary, that is, optimal values of synaptic weights aren't variable in time. Since in many practical tasks it is not so, for example, adaptive identification of non-stationary objects or non-stationary time series prediction, then high-performance adaptive learning algorithm having tracking and filtering properties can be used [Bodyanskiy, Kolodyazhniy, Stephan, 2001]:

$$\begin{cases} w_l^{ij}(k) = w_l^{ij}(k-1) + \eta_w(k)(y(k) - (w_l^{ij}(k-1))^{\mathrm{T}}\varphi^{ij}(k))\varphi^{ij}(k) = \\ = w_l^{ij}(k-1) + \eta_w(k)e_l(k)\varphi^{ij}(k), \\ \eta_w^{-1}(k) = r_w(k) = \alpha r_w(k-1) + \left\|\varphi^{ij}(k)\right\|^2, \quad 0 \le \alpha \le 1, \end{cases}$$

where α – smoothing parameter that defines compromise between filtering and tracking properties.

For the purpose of compartmental R-neuron approximating properties improving not only synaptic weights but also centers with receptive fields can be adjusted. At the same owing to Epanechnikov's kernels using learning algorithms have sufficiently simple form.

By using gradient procedure of criterion (2) minimization and its rate optimization technique [Otto, Bodyanskiy, Kolodyazhniy, 2003] we obtain compartmental R-neuron learning algorithm in the form

$$\begin{cases} w_l^{ij}(k) = w_l^{ij}(k-1) + \eta_w(k)e_l(k)\varphi^{ij}(k), \\ \eta_w^{-1}(k) = r_w(k) = \alpha r_w(k-1) + \left\|\varphi^{ij}(k)\right\|^2, \\ c_h^{ij}(k) = c_h^{ij}(k-1) + \eta_c(k)e_l(k)w_l^{ij}(k)\left(\Sigma_h^{ij}(k-1)\right)^{-1}(x^{ij}(k) - c_h^{ij}(k-1)) = \\ = c_h^{ij}(k-1) + \eta_c(k)e_l(k)g_h(k), \\ \eta_c^{-1}(k) = r_c(k) = \alpha r_c(k-1) + \left\|g_h(k)\right\|^2, \\ \left(\Sigma_h^{ij}(k)\right)^{-1} = \left(\Sigma_h^{ij}(k-1)\right)^{-1} - \eta_\Sigma(k)e_l(k)w_l^{ij}(k)(x^{ij}(k) - c_h^{ij}(k))(x^{ij}(k) - c_h^{ij}(k))^{\mathrm{T}} = \\ = \left(\Sigma_h^{ij}(k-1)\right)^{-1} - \eta_\Sigma(k)e_l(k)G_h(k), \\ \eta_\Sigma^{-1}(k) = \Gamma_\Sigma(k) = \alpha\Gamma_\Sigma(k-1) + Tr(G_h(k)G_h^{\mathrm{T}}(k)). \end{cases}$$

Evolving Cascade Neural Network

Uniting of GMDH and cascade neural networks ideas leads to architecture presented on fig. 2.

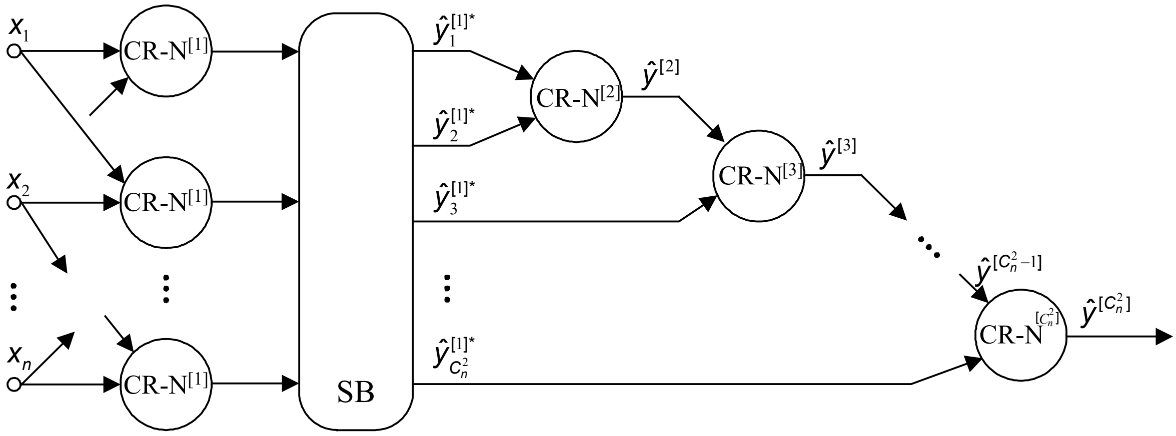


Fig. 2 – Evolving cascade neural network

The first hidden layer of the network is formed similarly to the first hidden layer of GMDH neural network [Pham, Liu, 1995] and contains the number of neurons equal to quantity of combinations of n in 2, that is C_n^2. Selection block SB executes sorting by accuracy, for example, in the sense of variations, of all output signals $\hat{y}_l^{[1]}$ so that the most accurate signal is $\hat{y}_1^{[1]*}$, then $\hat{y}_2^{[1]*}$ and the worst is $\hat{y}_{C_n^2}^{[1]*}$. Outputs of SB $\hat{y}_1^{[1]*}$ and $\hat{y}_2^{[1]*}$ then are fed to the one neuron of the second layer-cascade CR-N[2] that computes signal $\hat{y}^{[2]}$ which in the third cascade is combined with $\hat{y}_3^{[1]*}$. The process of cascades increasing lasts till required accuracy obtaining, at the same time, maximal neurons number of this network is restricted by the value $2C_n^2-1$. Thus, neural network is able to process information that are fed in real time by readjusting both its parameters and its architecture in time [Kasabov, 2003] and by adapting to the conditions of the solving task.

Conclusion

Architecture of evolving cascade radial-basis neural network was proposed in this paper. It is formed based on the idea of combining GMDH and cascade networks. Also this network is not disposed to the "curse of dimensionality" and is able to process information in real time by adapting its parameters and structure to the solving task conditions. Using of multidimensional Epanechnikov's kernels as activation functions allowed to introduce numerically simple learning algorithms that are characterized by high-performance.

Acknowledgements

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 the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Bibliography

[Ivakhnenko, 1969] Ivakhnenko A.G. Self-learning systems for recognition and automatic control. Technika, Kiev, 1969 (in Russian).

[Ivakhnenko, 1975] Ivakhnenko A.G. Long-term prediction and complex systems control. Technika, Kiev, 1975 (in Russian).

[Ivakhnenko, Stepashko, 1975] Ivakhnenko A.G., Stepashko V.S. Modeling noise-immunity. Naukova Dumka, Kiev, 1975 (in Russian).

[Ivakhnenko, Madala, 1994] Ivakhnenko A.G., Madala H.R. Inductive Learning Algorithms for Complex Systems Modeling. CRC Press, London-Tokio,1994.

[Pham, Liu, 1995] Pham D.J., Liu X. Neural Networks for Identification, Prediction and Control. Springer-Verlag, London, 1995.

[Avedyan, Barkan, Levin, 1999] Avedyan E. D., Barkan G.V., Levin I.K. Cascade neural networks // Avtomatika i Telemekhanika. – 1999. – №3. – P. 38-55. (in Russian).

[Epanechnikov, 1969] Epanechnikov V.A. Non-parametric estimate of multidimensional probability density // Probability theory and its application. – 1969. – 14. – №1. – P. 156-161. (in Russian).

[Friedman, Hastie, Tibshirani, 2003] Friedman J., Hastie T., Tibshirani R. The Elements of Statistical Learning. Data Mining, Inference and Prediction. Springer, Berlin, 2003.

[Bodyanskiy, Chaplanov, Kolodyazhniy, Otto, 2002] Bodyanskiy Ye., Chaplanov O., Kolodyazhniy V., Otto P. Adaptive quadratic radial basis function network for time series forecasting // Proc. East West Fuzzy Coll. 2002.– Zittau/Goerlitz: HS, 2002.– P.164-172.

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

[Otto, Bodyanskiy, Kolodyazhniy, 2003] Otto P., Bodyanskiy Ye., Kolodyazhniy V. A new learning algorithm for forecasting neuro-fuzzy network// Integrated Computer-Aided Engineering. – 2003. – 10. – №4. – P. 399-409.

[Kasabov, 2003] Kasabov N. Evolving Connectionist Systems: Methods and Applications in Bioinformatics. Springer–Verlag, London, 2003.

Authors' Information

***Yevgeniy Bodyanskiy** – Professor, Dr.-Ing. habil. Scientific head of Control Systems Research Laboratory, KhNURE, Senior member of IEEE, professor of Artificial Intelligence Department of KhNURE.*

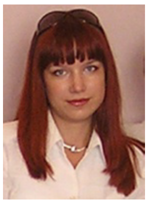
e-mail: bodya@kture.kharkov.ua

Major Fields of Scientific Research: hybrid systems of computational intelligence.

***Paul Grimm** – Professor, Computer Graphics Department, Erfurt University of Applied Sciences.*

e-mail: grimm@fh-erfurt.de

Major Fields of Scientific Research: image processing using intelligent computer technologies.



***Nataliya Teslenko** – Ph.D., Senior researcher of Control Systems Research Laboratory, KhNURE.*

e-mail: ntntp@ukr.net

Major Fields of Scientific Research: self-learning evolving neuro-fuzzy models and systems in the intelligent data analysis tasks