Hybrid radial-basis neuro-fuzzy wavelon in the non-stationary sequences forecasting problems

Bodyanskiy Yevgeniy, Vynokurova Olena

Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, av. Lenina 14, Kharkiv, 61166, Ukraine

bodya@kture.kharkov.ua, vinokurova@kture.kharkov.ua

Abstract. *Architecture of hybrid radial-basis neuro-fuzzy wavelon with adaptive membership-acivation function is considered. The learning algorithm for the all parameters of hybrid wavelon, providing the improvement of approximating properties that it check out the results of experimental simulation is proposed. This hybrid vavelon can be used as the node in the group method of data handling (GMDH) neural networks instead of the nonlinear adaline.*

Keywords

Artificial neural networks, computational intelligence, radial-basis neuro-fuzzy wavelon, wavelet theory, GMDH neural networks.

1 Introduction

At present time the neuro-fuzzy systems have been an increasingly popular technique of soft computing successfully applied for the processing of information containing complex nonlinear regularities and distortions of different kinds.

Alternative to the conventional and most wide spread multilayer ANNs is using the radial basis function networks (RBFN), which have one hidden layer consisting of so-called R-neurons. These networks learning is implemented on the output layer level which is usually represented by the adaptive linear associator [1, 2, 3, 4]. Unlike perceptron type of neurons (P-neurons), R-neurons usually 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.

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 centres allocation is remaining, and being unsuccessfully solved it 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.

Nowadays new class of ANNs – the wavelet neural networks are wide spread in the non-stationary signal processing under the uncertainty conditions. Elementary nodes of the wavelet neural networks are radial wavelons [5], where the activation functions are the even wavelets with argument in form the Euclidean distance between $x(k)$ and wavelet translation vector c_j, where every component of distance $\|x(k) - c_j\|$ is weighted by the dilation parameter. 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 function network and fuzzy inference systems [6, 7], and also possibility of using even wavelets as a membership functions [8, 9], within the bounds of the unification paradigm [5] we can talk about hybrid radial-basis neuro-fuzzy wavelon [11, 12, 13, 14, 15] having the radial-basis function network fast learning ability, fuzzy inferences systems interpretability and wavelet's local properties.

2 Hybrid radial-basis neuro-fuzzy wavelon

Let us consider the two-layers hybrid system. The input layer is the receptive and in current time instant k the input signal in vector form $x(k) = (x_1(k), x_2(2), \ldots, x_n(k))^T$ is fed on it. Unlike radial basis function network the hidden layer consists not of R-neurons, but by wavelons with wavelet type activation function in the form

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

in which instead of translation parameters σ_{ji} the positive-definite dilation matrix Q_j is used. It moves us from using Euclidean distance to implementation of Itakura-Saito metric [10].

This results in the fact that receptive fields – hyperellipsoids can have the arbitrary orientation relatively to the coordinate axes of space X, what extends the functional properties of radial-basis neuro-fuzzy wavelon.

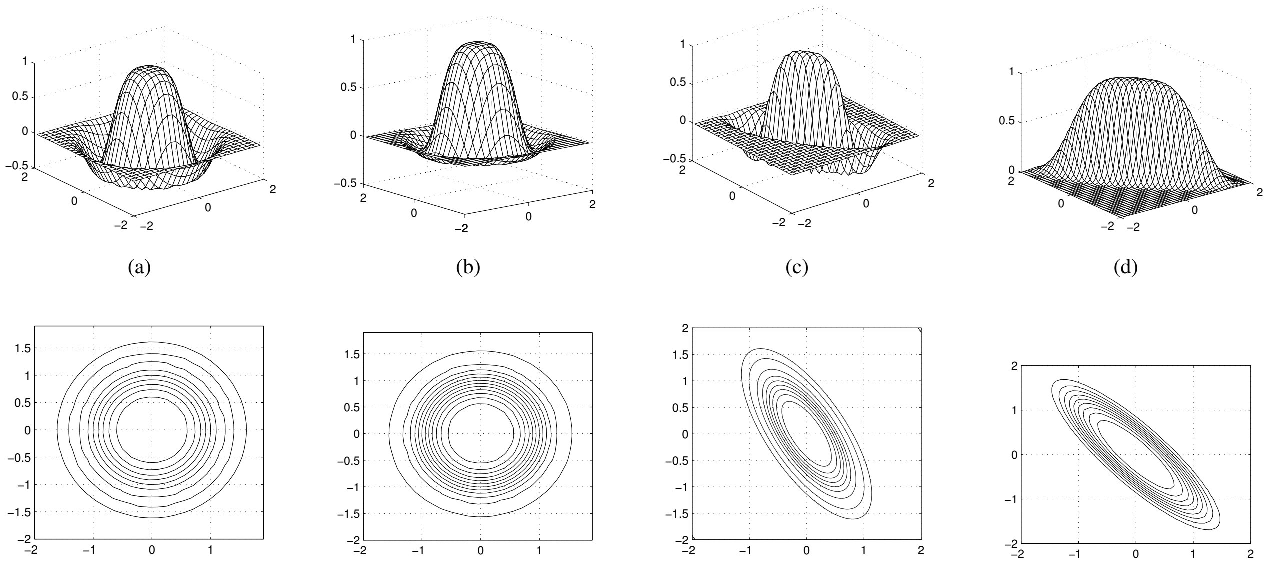
Based on the results about that the wavelet-function can be used as a membership function in fuzzy systems, we can introduce the adaptive membership-activation function based on wavelet Mexican Hat, having form

$$\varphi_j(x(k)) = (1 - \alpha_j \tau_j^2(x(k))) \exp(-\tau_j^2(x(k))/2), \tag{2}$$

where $\tau_j(x(k)) = ((x(k) - c_j(k))^T Q_j^{-1}(k)(x(k) - c_j(k)))$, α_j is tuning parameter ($0 \leq \alpha_j \leq 1$).

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

Fig.1 shows the wavelon membership-activation function (2) with arbitrary matrices Q_j and different parameters α_j.



(a) (b) (c) (d)

(e) $Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \alpha_1 = 1$ (f) $Q_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \alpha_2 = 0.5$ (g) $Q_3 = \begin{pmatrix} 4 & 3 \\ 1 & 2 \end{pmatrix}, \alpha_3 = 1$ (h) $Q_4 = \begin{pmatrix} 4 & 4 \\ 2 & 3 \end{pmatrix}, \alpha_4 = 0$

Fig. 1. Wavelons membership-activation function with arbitrary matrices Q_j and variable parameters α_j

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

$$\hat{y}(k) = w_0 + \sum_{j=1}^{h} w_j \varphi((x(k) - c_j)^T Q_j^{-1}(x(k) - c_j)) = w^T \varphi(x(k)), \tag{3}$$

where $\varphi_0(x(k)) \equiv 0$, $w = (w_0, w_1, \ldots, w_h)^T$, $\varphi(x(k)) = (1, \varphi_1(x(k)), \varphi_2(x(k)), \ldots, \varphi_h(x(k)))^T$.

3 Learning algorithm for hybrid radial-basis neuro-fuzzy wavelon

For the synaptic weights w_j and the wavelon parameters (vectors c_j, matrices Q_j and parameters α_j) tuning we use gradient minimization of criterion

$$E(k)=\frac{1}{2}e^2(k)=\frac{1}{2}(y(k)-\hat{y}(k))^2, \tag{4}$$

so unlike in the component-wise learning considered in [4], we make some corrections in the vector-matrix form, that, firstly is less computationally expensive, and secondly it allows to optimize learning process on the operation rate.

In general case the learning algorithm can be written in form

$$\begin{cases} w_j(k+1)=w_j(k)-\eta_w\left(\partial E(k)/\partial w_j\right), j=1\dots,h,\\ c_j(k+1)=c_j(k)-\eta_c\nabla_{c_j}E(k), j=1,\dots,h,\\ Q_j^{-1}(k+1)=Q_j^{-1}(k)-\eta_Q\left\{\dfrac{\partial E(k)}{\partial Q_j^{-1}}\right\}, j=1,\dots,h,\\ \alpha_j(k+1)=\alpha_j(k)-\eta_\alpha\left(\partial E(k)/\partial\alpha_j\right), j=1,\dots,h, \end{cases} \tag{5}$$

where $\nabla_{c_j}E(k)$ is $n\times 1$-vector-gradient criterion (4) on c_j; $\left\{\dfrac{\partial E(k)}{\partial Q_j^{-1}}\right\}$ is $(n\times n)$-matrix, formed by partial derivatives $E(k)$ on components Q_j^{-1}; η_w, η_{c_j}, $\eta_{Q_j^{-1}}$ and η_{α_j} are the learning rates.

For the adaptive membership function (4) we can write

$$\begin{cases} \partial E(k)/\partial w_j=e(k)\left(1-\alpha_j\tau_j^2(x(k))\right)\exp(-\tau_j^2(x(k))/2)=e(k)J_{w_j}(k),\\ \nabla_{c_j}E(k)=2e(k)w_j(k)\left(\alpha_j\tau_j^3(x(k))-(2\alpha_j+1)\tau_j(x(k))\right)\exp(-\tau_j^2(x(k))/2)Q_j^{-1}(x(k)-c_j(k))=\\ \qquad =e(k)J_{c_j}(k),\\ \left\{\dfrac{\partial E(k)}{\partial Q_j^{-1}}\right\}=e(k)w_j(k)\left(\alpha_j\tau_j^3(x(k))-(2\alpha_j+1)\tau_j(x(k))\right)\exp(-\tau_j^2(x(k))/2)(x(k)-c_j(k))\cdot\\ \qquad \cdot(x(k)-c_j(k))^T=-e(k)J_{Q_j^{-1}}(k),\\ \partial E(k)/\partial\alpha_j=-e(k)w_j(k)\tau_j^2(x(k))\exp(-\tau_j^2(x(k))/2)=e(k)J_{\alpha_j}(k), \end{cases} \tag{6}$$

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

Increasing of the learning rate can be achieved by using more complex procedures than gradient ones, such as Hartley or Marquardt procedures and using the inverse matrices lemma and after applying simple transformations we obtain the effective parameters learning algorithm in the form

$$w_j(k+1)=w_j(k)+\lambda_w\left(e(k)J_w(k)/(\eta_w+J_w^2(k))\right), \tag{7}$$

$$c_j(k+1)=c_j(k)-\lambda_c\left(e(k)J_c(k)/(\eta_c+\|J_c(k)\|^2)\right), \tag{8}$$

$$\alpha_j(k+1)=\alpha_j(k)-\lambda_\alpha\left(e(k)J_\alpha(k)/(\eta_\alpha+J_\alpha^2(k))\right), \tag{9}$$

accurate within the descriptions and coinciding with optimal (for $\lambda_w=\lambda_c=\lambda_\alpha=1$, $\eta_w=\eta_c=\eta_\alpha=0$) one-step Kaczmarz algorithm.

For the turning of matrices Q_j^{-1} we can use the matrix modification of algorithm (8) [14] in form

$$Q_j^{-1}(k+1)=Q_j^{-1}(k)+\lambda_Q\left(e(k)J_Q(k)/\left(\eta_Q+Tr\left(J_Q^T(k)J_Q(k)\right)\right)\right), \tag{10}$$

where λ_Q is a positive dampening parameter and η_Q is a momentum term parameter.

It is known, that one-step algorithms such as Kaczmarz one, have rapid response, but they don't have filtering properties, i.e. they are not operated well in the conditions of intensive disturbances and noises. In order to provide the learning algorithm with smoothing properties, we can introduce learning algorithm:

$$\begin{cases} w_j(k+1)=w_j(k)+\lambda_w\dfrac{e(k)J_w(k)}{\eta_w(k)}, \quad \eta_w(k+1)=\gamma_w\eta_w(k)+J_w^2(k+1),\\ c_j(k+1)=c_j(k)-\lambda_c\dfrac{e(k)J_c(k)}{\eta_c(k)}, \quad \eta_c(k+1)=\gamma_c\eta_c(k)+\|J_c(k+1)\|^2,\\ Q_j^{-1}(k+1)=Q_j^{-1}(k)+\lambda_Q\dfrac{e(k)J_Q(k)}{\eta_Q(k)}, \quad \eta_Q(k+1)=\gamma_Q\eta_Q(k)+Tr\left(J_Q^T(k)J_Q(k)\right),\\ \alpha_j(k+1)=\alpha_j(k)+\lambda_\alpha\dfrac{e(k)J_\alpha(k)}{\eta_\alpha(k)}, \quad \eta_\alpha(k+1)=\gamma_\alpha\eta_\alpha(k)+J_\alpha^2(k+1), \end{cases} \tag{11}$$

(here $0 \leq \{\lambda_w, \lambda_c, \lambda_Q, \lambda_\alpha\} \leq 1$ are parameters of weighted out-dated information). This procedure is nonlinear hybrid of Kaczmarz-Widrow-Hoff and Goodwin-Ramadge-Caines algorithms and has both following and filtering properties.

4 Conclusion

In the paper computationally simple and significantly effective learning algorithm with radial-basis neuro-fuzzy wavelon parameters and adaptive wavelet membership-activation function parameter is proposed. It combines both following and filtering properties and allows processing of non-stationary nonlinearly signals in real time. Using the wavelons receptor fields, including their transformations (translation, dilation, rotation, transformation membership-activation function form) allows to improve the network approximation properties, what is confirmed by the experiments research results. In the future the proposed adaptive hybrid radial-basis neuro-fuzzy wavelon will be used as the node in the multilayer GMDH neural networks instead of conventional non-linear adaline.

References

[1] Moody J., Darken C. J.: Fast learning in networks of locally-tuned processing units. *N*eural Computation, 1: 281-294, 1989.

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

[3] Park J., Sandberg I. W.: Universal approximation using radial-basis-function networks. *N*eural Computation, 3: 246-257, 1991.

[4] Bishop C. M.: Neural Networks for Pattern Recognition. *O*xford: Clarendon Press, 1995.

[5] Reyneri L. M.: Unification of neural and wavelet networks and fuzzy systems. *I*EEE Trans. on Neural Networks, 10: 801-814, 1999.

[6] JangSun:1993,HuntHaasSmith:1996 Jang J. S. R., Sun C. T.: Functional equivalence between radial basis function networks and fuzzy inference systems. *I*EEE Trans. on Neural Networks, 4: 156-159, 1993.

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

[8] Mitaim S., Kosko B.: What is the best shape for a fuzzy set in function approximation? *I*n proceedings of the 5th IEEE Int. Conf on Fuzzy Systems "Fuzz-96", 2: 1237-1213, 1996.

[9] Mitaim S., Kosko B.: Adaptive joint fuzzy sets for function approximation. *I*n proceedings of the Int. Conf. on Neural Networks "ICNN-97", 1997.

[10] Itakura F.: Maximum prediction residual principle applied to speech recognition. *I*EEE Trans. on Acoustics, Speech and Signal Processing, 23: 67-72, 1975.

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

[12] Bodyanskiy Ye., Pliss I., Vynokurova O.: A learning algorithm for forecasting adaptive wavelet-neuro-fuzzy network. *I*n Proceedings of the XIII -th Int. Conf. "Information Research and Applications", Varna (Bulgaria), 211-218, 2007.

[13] Bodyanskiy Ye., Bednarska G., Vynokurova O.: Forecasting adaptive wavelet-neuro-fuzzy-network using Polywog activation function. *I*n Proceedings of the 3-rd International Conference "Advanced Computer Systems and Networks: Design and Application", Lviv, 79-82, 2007.

[14] Bodyanskiy Ye., Pliss I., Vynokurova O.: Adaptive wavelet-neuro-fuzzy network in the forecasting and emulation tasks. *I*nt. Journal on Information Theory and Applications, 15 (1): 47-55, 2008.

[15] Bodyanskiy Ye., Pavlov O., Vynokurova O.: Outliers resistant learning algorithm for radial-basis-fuzzy-wavelet-neural network in stomach acute injury diagnosis tasks. *E*ds. By K.Markov, K.Ivanova, I.Mitov International Book Series "Information Science and Computing", Sofia: Institute of Information Theories and Application FOI ITHEA, 2: 55-62, 2008.