UDC 519.7:007.52

Ye. Bodyanskiy, O.Vynokurova

ROBUST LEARNING ALGORITHM FOR WAVELET-NEURAL-FUZZY NETWORK BASED ON POLYWOG WAVELET

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, intelligence 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 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. 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 ANN are the radial basis function networks (RBFN), having one hidden layer [1-7]. The principal advantage of RBFN is the high learning rate in the output layer, because the turning parameters are linearly included to the network description. At the same time the problem of neurons centers allocation is remaining, and its unsuccessful solving leads to the «curse of dimensionality» problem.

Along with neural networks for the arbitrary type signals processing, in the last years the wavelet theory is used sufficiently often [8], 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 [9-12] have evolved their efficiency for the analysis of nonstationary nonlinear signals and processes. Elementary nodes of the wavelet neural networks are so-called radial wavelons. 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 [13], and also possibility of using even wavelet as a membership function [14], within the bounds of the unification paradigm [15] we can talk about such hybrid system as Radial-Basis-Function-Wavelet-Neuro-Fuzzy Network (RBFWNFN) having the radial-basis function network fast learning ability, fuzzy inferences systems interpretability and wavelet's local properties.

Mostly tuning algorithms based on square learning criteria and 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 that allow signal processing in presence of anomalous outliers.

This paper is devoted to synthesis of robust learning algorithm RBFWNN, 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.

1. Radial-basis-fuzzy-wavelet-neural-network architecture

Let us consider the two-layers architecture 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 form

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

in which the positive-definite dilation matrix Q_j is used, i.e. it is Itakura-Saito metric. This results to the fact that receptive field − wavelons hyperellipsoids can have the arbitrary orientation relatively to the coordinate axes of space X, what extends the functional properties of RBFWNN.

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

$$\hat{y}(k)=w_0+\sum_{j=1}^{h} w_j\varphi\left((x(k)-c_j)^T Q_j^{-1}(x(k)-c_j)\right)=w^T\varphi(x(k)), \qquad \textbf{(2)}$$

where $\varphi_0(x(k)) \equiv 1$, $w = (w_0, w_1, w_2, \ldots, w_h)^T$, $\varphi(x(k)) = (1, \varphi_1(x(k)), \varphi_2(x(k)), \ldots, \varphi_h(x(k)))^T$.

2. The robust learning algorithm for RBFWNFN

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 outliers, an outrage errors, and non-Gaussian disturbance with “heavy tails” the methods based on the least square criterion loose their efficiency. In this case the methods of robust estimation [16] which can be used too for the learning of the artificial neural networks [17, 18] appear on the first role.

Introducing into the consideration the learning error

$$e(k) = y(k) - \hat{y}(k) = y(k) - w^T(k)\varphi(k) \tag{3}$$

and robust identification criterion by R. Welsh [19, 20]

$$E(k) = f(k) = \beta^2 \ln\left(\cosh\left(e(k)/\beta\right)\right), \tag{4}$$

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 (4) satisfies to the metric space axioms.

Further we shall consider synthesis of the learning algorithms. For the synaptic weights and the waveleon parameters (vectors c_j and matrices Q_j^{-1}) tuning we use gradient minimization of criterion (4), thus unlike the component-wise learning considered in [7], we shall make some correction in the vector-matrix form, that, firstly is easier from computing point of view, and secondly it will allow to optimize learning process on the operation rate.

For arbitrary wavelet $\varphi((x(k)-c_j)^T Q_j^{-1}(x(k)-c_j))$ we can write

$$\begin{cases} \nabla_w E(k) = -\beta \tanh\left(e(k)/\beta\right)\varphi_j((x(k)-c_j(k))^T Q_j^{-1}(k)(x(k)-c_j(k))) = \\ = -\tanh\left(e(k)/\beta\right) J_w(k), \\ \nabla_{c_j} E(k) = \beta \tanh\left(e(k)/\beta\right) w_j(k)\varphi_j'((x(k)-c_j(k))^T Q_j^{-1}(k)(x(k)-c_j(k)))\cdot \\ \cdot Q_j^{-1}(k)(x(k)-c_j(k)) = \tanh\left(e(k)/\beta\right) J_{c_j}(k), \\ \left\{\partial E(k)\big/\partial Q_j^{-1}\right\} = -\beta \tanh\left(e(k)/\beta\right) w_j(k)\varphi_j'\left((x(k)-c_j(k))^T Q_j^{-1}(k)\cdot\right. \\ \left.\cdot(x(k)-c_j(k))\right)(x(k)-c_j(k))(x(k)-c_j(k))^T = -\tanh\left(e(k)/\beta\right) J_{Q_j^{-1}}(k), \end{cases} \tag{5}$$

where $\nabla_w E$ is vector-gradient of the criterion **(4)** on w, $\nabla_{c_j} E$ is $(n\times 1)$-vector-gradient on c_j; $\left\{\partial E(k)\big/\partial Q_j^{-1}\right\}$ is $(n\times n)$-matrix, formed the partial derivatives E on components Q_j^{-1}; η_w, η_{c_j}, $\eta_{Q_j^{-1}}$ are the learning rates; $\varphi_j'(\bullet)$ is the derivative j-th wavelet on the argument $(x(k)-c_j(k))^T Q_j^{-1}(k)(x(k)-c_j(k))$.

Then the wavelons learning algorithm of the hidden layer subject to **(5)** is taking the form

$$\begin{cases} w(k+1) = w(k) + \eta_w \beta \tanh\left(e(k)/\beta\right)\varphi_j((x(k)-c_j(k))^T Q_j^{-1}(k)(x(k)-c_j(k))) = \\ = w(k) + \eta_w \tanh\left(e(k)/\beta\right) J_w(k), \\ c_j(k+1) = c_j(k) - \eta_{c_j} \beta \tanh\left(e(k)/\beta\right) w_j(k)\varphi_j'((x(k)-c_j(k))^T Q_j^{-1}(k)(x(k)-c_j(k)))\cdot \\ \cdot Q_j^{-1}(k)(x(k)-c_j(k)) = c_j(k) - \eta_{c_j} \tanh\left(e(k)/\beta\right) J_{c_j}(k), \\ Q_j^{-1}(k+1) = Q_j^{-1}(k) + \eta_{Q_j^{-1}} \beta \tanh\left(e(k)/\beta\right) w_j(k)\varphi_j'((x(k)-c_j(k))^T Q_j^{-1}(k)\cdot \\ \cdot(x(k)-c_j(k)))(x(k)-c_j(k))(x(k)-c_j(k))^T = Q_j^{-1}(k) + \eta_{Q_j^{-1}} \tanh\left(e(k)/\beta\right) J_{Q_j^{-1}}(k). \end{cases} \tag{6}$$

For the Polywog wavelet **[21]** we can rewrite the algorithm **(6)** in the relatively simple form

$$\begin{cases} w(k+1) = w(k) + \eta_w \beta \tanh\left(e(k)/\beta\right)\left(1-\Psi^2\right)\exp\left(-\Psi^2/2\right), \\ c_j(k+1) = c_j(k) - \eta_{c_j} \beta \tanh\left(e(k)/\beta\right) w_j(k)\Psi\left(\Psi^2-3\right)\exp\left(-\Psi^2/2\right)\cdot \\ \cdot Q_j^{-1}(k)(x(k)-c_j(k)), \\ Q_j^{-1}(k+1) = Q_j^{-1}(k) + \eta_{Q_j^{-1}} \beta \tanh\left(e(k)/\beta\right) w_j(k)\Psi\left(\Psi^2-3\right)\exp\left(-\Psi^2/2\right)\cdot \\ \cdot(x(k)-c_j(k))(x(k)-c_j(k))^T, \end{cases} \tag{7}$$

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

Conclusion

In the paper computationally simple and effective robust learning algorithm of all RBFWNN parameters is proposed. This learning algorithm allows on-line processing of non-linear signals under a number of outliers and "heavy tails" disturbations. Addition of wavelons receptive fields, including their transformations (dilation, translation, rotation) permits improve the network approximation properties, that is confirmed by the experiments research results.

BIBLIOGRAPHY

1. Moody J. Darken C. J. Fast learning in networks of locally-tuned processing units // Neural Computation. – 1989. –1. – P. 281-294.
2. Moody J., Darken C. 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.
3. Park J., Sandberg I. W. Universal approximation using radial-basis-function networks // Neural Computation. – 1991. – 3. – P. 246-257.
4. Leonard J. A., Kramer M. A., Ungar L. H. Using radial basis functions to approximate a function and its error bounds // IEEE Trans. on Neural Networks. – 1992. – 3. – P. 614-627.
5. Sunil E. V. T., Yung C. Sh. Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems // IEEE Trans. on Neural Networks. – 1994. – 5. – P. 594-603.
6. Poggio T., Girosi F. A Theory of Networks for Approximation and Learning. – A. I. Memo № 1140, C.B.I.P. Paper № 31, Massachussetts Institute of Technology. – 1994. – 63 p.
7. Bishop C. M. Neural Networks for Pattern Recognition. - Oxford: Clarendon Press, 1995. – 482 p.
8. Daubechies I. Ten Lectures on Wavelets. - Philadelphia, PA: SIAM., 1992. – 228 p.
9. Billings S. A., Wei H.-L. A new class of wavelet networks for nonlinear system identification // IEEE Trans. on Neural Networks. – 2005. – 16. - P. 862-874.
10. Bodyanskiy Ye., Lamonova N., Vynokurova O. Recurrent learning algorithm for double-wavelet neuron // Proc. XII -th Int. Conf. "Knowledge - Dialogue - Solution". - Varna (Bulgaria). – 2006. - P.77-84.
11. Bodyanskiy Ye., Lamonova N., Vynokurova O. Double-wavelet neuron based on analytical activation functions // Int. J. Information Theory and Applications. – 2007. – 14. – P. 281-288.
12. Bodyanskiy Ye., Pliss I., Vynokurova O. A learning algorithm for forecasting adaptive wavelet-neuro-fuzzy network // Proc. XIII -th Int. Conf. "Information Research & Applications". – Varna (Bulgaria). – 2007. – P.211-218.

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

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

15. Reyneri L. M. Unification of neural and wavelet networks and fuzzy systems // IEEE Trans. on Neural Networks. – 1999. – 10. – P. 801-814.

16. Rey W.J.J. Robust Statistical Methods // Lecture Notes in Mathematics. – 2. – Berlin-Heidelberg-New York: Springer-Verlag, 1978. – 128 p.

17. Cichocki A., Unbehauen R. Neural Networks for Optimization and Signal Processing. – Stuttgart: Teubner, 1993. – 526 p.

18. Sheng-Tun Li, Shu-Ching Chen Function approximation using robust wavelet neural networks // Proc. of the 14th IEEE Int. Conf. on Tools with Artificial Intelligence. – 2002. – P. 483- 488.

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

20. Welsh R.E. 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.

21. Lekutai G., van Landingham H. F. Self-tuning control of nonlinear systems using neural network adaptive frame wavelets // *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics.* - Piscataway, N.J. – 1997. – 2. - P. 1017-1022.

Получено 11.03.2008 г.