MULTILAYER NEURO-FUZZY SYSTEM FOR SOLVING ON-LINE DIAGNOSTICS TASKS

Yevgeniy Bodyanskiy, Olena Vynokurova, Iryna Pliss, Dmytro Peleshko

***Abstract**: In the paper the problem of on-line diagnostics and properties change detection of systems whose output signal is multidimensional non-stationary stochastic sequence is considered. The six-layer diagnostic neuro-fuzzy system is proposed. The first layer of this system consists of membership functions blocks, the second layer provides aggregation of fuzzyfied inputs, the third one consists of tuning synaptic weights, the fourth one consists of summing blocks set, the fifth layer produces normalization (defuzzification) of output signals and finally sixth layer consists of nonlinear activation functions and provides the properties change detection. So the proposed neuro-fuzzy system is modification of L.Wang-J.Mendel system and solves diagnostics-classification problems in real time mode. For tuning synaptic weights of proposed system we have used special learning criterion, that is aimed at solving of pattern recognition, classification, diagnostics problems etc. The learning algorithm for synaptic weights is proposed and its speed optimization is performed. It allowed to design recurrent procedure, which is matrix hybrid of J.Shynk and S. Kaczmarz-B.Widrow-M.Hoff learning algorithm. It is important to notice that proposed system has significantly fewer number of tuning synaptic weights in comparison with conventional well-known neural network diagnosis systems based on multilayer perceptrons or radial basis function networks. This feature allows to reduce the learning set volume, to achieve optimal training rate, to provide linguistic interpretability and «transparency» of obtained results.*

***Keywords**: neuro-fuzzy-system, adaptive learning, data mining, diagnostic*

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

Introduction

For solving wide class problems of Data Mining which connected first of all with diagnostic, classification, pattern recognition etc the artificial neural networks are used increasingly frequently due to their universal approximation properties and learning ability based on experimental data set. Although for solving such problem the conventional multilayer perceptron is used in most cases, it should not go unnoticed such its common disadvantages as sufficiently large training set volume, low convergence rate of backpropagation learning algorithm, the necessity for using a large number of training epoch. And if especially computational problems we can solve, but necessity for a representative training sample significantly complicates the use of this neural network for solving many practical problems. This problem appears especially in the research where data set has short dimension and at that the object is described by set of different characteristics [Swamy, 2014, Kurse, 2013].

In this situation radial basis function neural networks are preferable [Haykin, 1999], whose output signal is linearly dependent on tuning synaptic weights. This fact allows to use for training these networks large range of well-known approaches from conventional least squares method to the popular linear adaptive identification algorithms [Ljung, 1999]. And although specificity of diagnosis-classification problems restricts to use conventional square learning criterion, using special Shynk criterion [Shynk, 1990], focused on pattern recognition tasks with binary training signal allows to design sufficient simple and effective diagnostic radial basis function neural network [Bodyanskiy, 2002, Bodyanskiy, 2005].

In spite of all its advantages radial basis neural networks is not panacea for all cases because of its possibilities are limited by so called "curse of dimensionality" that leads to the exponential increasing number of tuning synaptic weights in accordance with input signal-pattern dimension space.

Overcome this problem, the procedure of preliminary setting of radial basis function centers by one or another clustering methods is used. Hence, supervised learning is completed by self-learning of its centers, what makes such learning too tedious.

In [Bodyanskiy, 2010] for solving tasks of text documents processing in the context of Text Mining, which is characterized by large dimensionality of input signals, hierarchical radial basis function network with multilayer architecture was proposed. Such system uses usual RBFN in each unit and on the input of systems only part of features is fed, what allows to overcome problem of "course dimensionality". The main thing that has been achieved in this situation - it is the possibility to operate under conditions when input pattern dimension is comparable with training set volume. At the same time, it should not go unnoticed inconvenience of this system, impossibility to operate in sequential on-line mode, high level of subjectivity in partition of input pattern into subvector-vector set for each network unit.

Anyway the problem of processing data with high dimension of features vector under condition, when training set volume is comparable with this dimension, is attractive and especially for solving tasks of classification and diagnostic in Text Mining, Web Mining and medical-biology applications. Using of neuro-fuzzy systems (NFS) [Jang, 1997] is significantly future-oriented because such systems allow to provide not only good approximation properties and learning ability, but linguistic interpretability of obtained results. It is also necessary to note that obtained results of NFS-systems are equivalent to the results of radial basis function networks [Jang, 1993], this fact allows to use identical learning algorithms.

Thus this paper is devoted to synthesis of diagnostic neuro-fuzzy system for the case, when training set dimension is comparable with input patterns set volume, and these patterns are fed for processing in on-line mode.

Diagnostic neuro-fuzzy system architecture

Architecture of considered NFS is shown on Fig. 1 and consists of six sequentially-connected layers. In the input (null, receptive) layer of NFS $(n\times 1)$-dimensional vector of input signals-patterns $x(k)=(x_1(k),x_2(k),\ldots x_n(k))^T$ is fed, where $k=1,2,\ldots,N$ is observation number in initial data set. In this case it is supposed that all components $x_i(k)$ preliminary are modified so that

$$0\le x_i(\mathrm{k})\le 1, \forall i=1,2,\ldots,n,$$

and binary input features have value 0 or 1.

The first layer consists of nh membership function $\mu_{li}(x_i(k))$, $i=1,2,\ldots,n$; $l=1,2,\ldots,h$ and provides fuzzyfication of input variables, at that the larger the number h, the better approximating properties NFS, although it is enough to have $h=2$ for binary features.

The second hidden layer realizes aggregation of membership levels, which are computed in the first layer, and consists of h multiplication units Π.

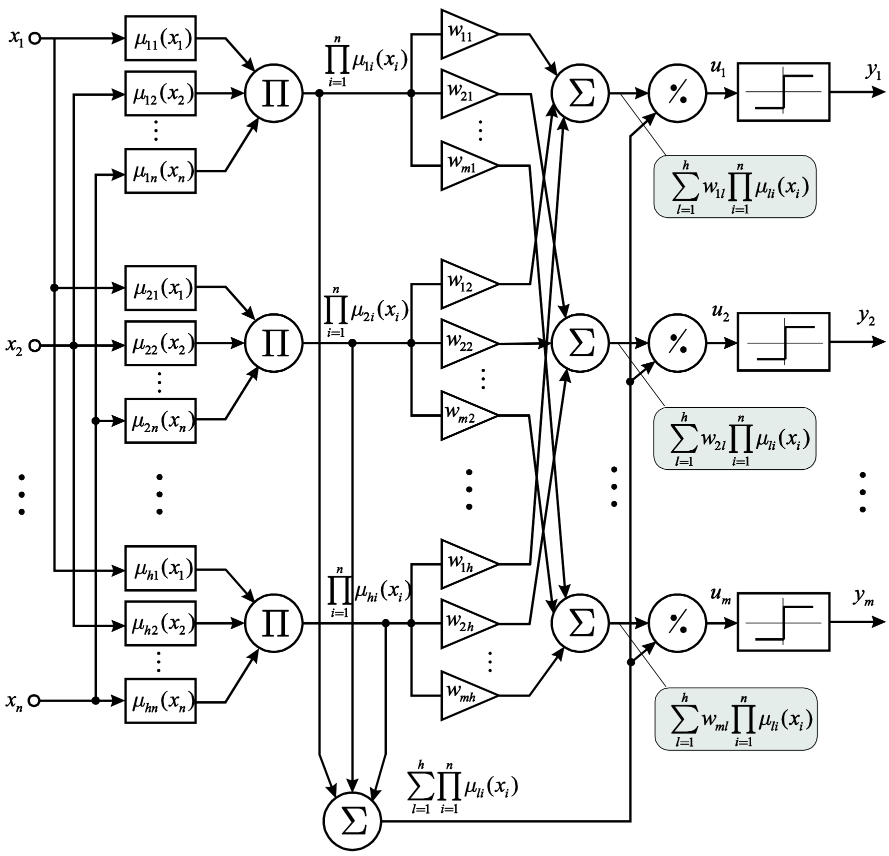


Fig. 1. Diagnostic neuro-fuzzy system

The third hidden layer is one of synaptic weights w_{jl}, $j = 1, 2, \ldots, m$ which are adjusted during learning process. Proposed NFS consists of mh tuning weights, where m is a number of potential classes, one for each system output. It is clear that $mh << e^n$, i.e. number of NFS weights are significantly smaller than the number of RBFN weights.

The fourth hidden layer consists of $m+1$ summators Σ, which compute sum of output signal of the second and the third hidden layers.

In fifth hidden layer that consists of m division unit / normalization of fourth layer output signals is realized.

And finally output (sixth) layer consists of m non-linear activation functions, at that in diagnosis tasks it is reasonable to use the simplest signum-functions, which takes $+1$ value in case of right diagnosis, and -1 – otherwise. Therefore output system signals $y_j(k)$ can take only two values ± 1.

Thus if vector signal $x(k)$ is fed on NFS input, the first layer elements compute membership levels $\mu_{li}(x_i(k))$, at that usually the bell-shaped (kernel) construction with as membership function nonstrictly local receptive field are used as membership functions. It allows to avoid appearing of “gaps” in fuzzyficated space [Friedman, 2003]. Most often it is conventional Gaussians.

$$\mu_{li}(x_i(k)) = \exp\left(-\frac{(x_i(k) - c_{li})^2}{2\sigma_i^2}\right) \qquad (1)$$

where c_{li} is center parameter (in the simplest case the centers are located uniformly in the interval [0,1] with step $(h-1)^{-1}$), σ_i is width parameter, selected empirically or tuning with backpropagation algorithm [Osowski, 2006]. Fig. 2 shows membership functions.

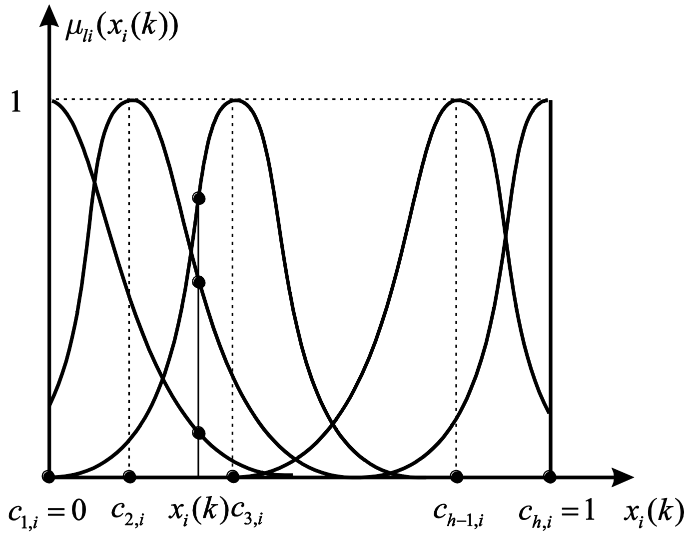


Fig. 2 – Bell-shaped membership functions

It is clear that for binary variables $x_i(k)$ it is enough to use only two triangular membership functions

$$\begin{cases} \mu_{1i}(x_i(k)) = 1 - x_i(k), \\ \mu_{2i}(x_i(k)) = x_i(k), \end{cases} \tag{2}$$

that are shown on Fig. 3.

We also have to notice that membership functions (2) in some cases with success can be used for features which have arbitrary number of values (see fig. 3), and number of synaptic weights take on minimally possible value $2m$.

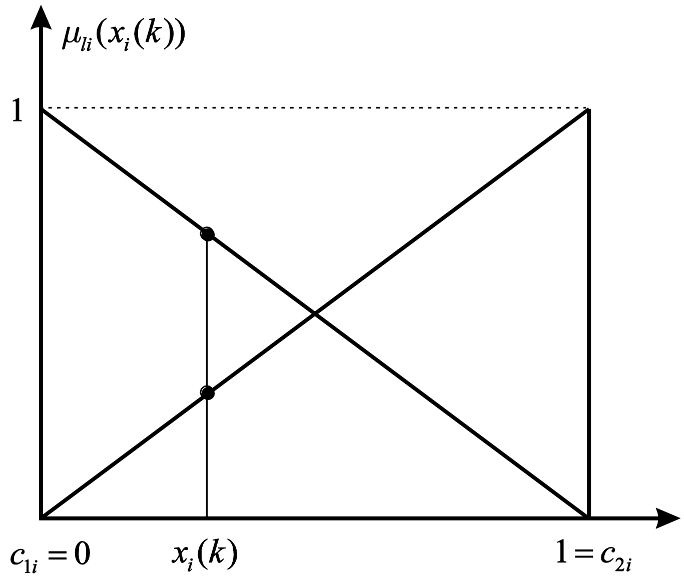


Fig. 3 – Membership functions for binary variables

On the outputs of second layer the aggregated values $\prod_{i=1}^{n}\mu_{li}(x_i(k))$ are appeared, at that it is simple to notice that if width parameters σ_i are the same for all features, i.e. $\sigma_i = \sigma$, that

$$\prod_{i=1}^{n}\mu_{li}(x_i(k)) = \prod_{i=1}^{n}\exp\left(-\frac{(x_i(k)-c_{li})^2}{2\sigma^2}\right) = \exp\left(-\frac{\|x_i(k)-c_{li}\|^2}{2\sigma^2}\right)$$

(here $c_l = (c_{l_1}, c_{l_2}, \ldots, c_{l_n})^T$) i.e. nonlinear transformation similar RBFN is realized.

Outputs of third hidden layer are values $w_{jl}\prod_{i=1}^{n}\mu_{li}(x_i(k))$, forth one $\sum_{l=1}^{h}w_{jl}\prod_{i=1}^{n}\mu_{li}(x_i(k))$ and $\sum_{l=1}^{h}\prod_{i=1}^{n}\mu_{li}(x_i(k))$, fifth one

$$u_j(k) = \frac{\sum_{l=1}^{h}w_{jl}\prod_{i=1}^{n}\mu_{li}(x_i(k))}{\sum_{l=1}^{h}\prod_{i=1}^{n}\mu_{li}(x_i(k))} = \sum_{l=1}^{h}w_{jl}\frac{\prod_{i=1}^{n}\mu_{li}(x_i(k))}{\sum_{l=1}^{h}\prod_{i=1}^{n}\mu_{li}(x_i(k))} =$$

$$= \sum_{l=1}^{h}w_{jl}\varphi_l(x(k)) = w_j^T\varphi(x(k))$$

(here $\varphi_l(x(k)) = \prod_{i=1}^{n}\mu_{li}(x_i(k))\left(\sum_{l=1}^{h}\prod_{i=1}^{n}\mu_{li}(x_i(k))\right)^{-1}$, $w_j = (w_{j1}, w_{j2}, \ldots, w_{jh})^T$, $\varphi(x(k)) = (\varphi_1(x(k)), \varphi_2(x(k)), \ldots, \varphi_h(x(k)))^T$) and, finally, sixth

$$y_j(k) = \operatorname{sign} u_j(k)$$

It is clearly to see that proposed NFS is modification of Wang-Mendel system [Wang, 1992, Wang, 1994], which oriented for solving on-line diagnostic-classification tasks.

Diagnostic neuro-fuzzy system learning

For training of synaptic weights on system under consideration we use learning algorithm based on specialized criterion [Shynk, 1990], which is aimed for solving pattern recognition, classification, diagnostic tasks etc.

Let us introduce m errors of learning

$$e_j(k) = d_j(k) - y_j(k) = d_j(k) - \operatorname{sign} u_j(k)$$

and m criterions based on these errors

$$\begin{aligned} E_j(k) &= e_j(k)u_j(k) = d_j(k)u_j(k) - |u_j(k)| = \\ &= \left(d_j(k) - \operatorname{sign} w_j^T\varphi(x(k))\right)\cdot w_j^T\varphi(x(k)), \end{aligned} \tag{3}$$

where $d_j(k) \in \{-1, 1\}$ is training signal, having value 1, if input vector $x(k)$ belongs to j –th diagnosis, and -1 otherwise.

For synaptic weights tuning we can use conventional gradient procedure of criterion minimization (3)

$$w_{jl}(k+1) = w_{jl}(k) - \eta(k)\frac{\partial E_j(k)}{\partial w_{jl}}$$

(here $\eta(k)$ is learning rate parameter), which on vector form can be rewritten in the form

$$\begin{aligned} w_j(k+1) &= w_j(k) + \eta(k)e_j(k)\varphi(x(k)) = \\ &= w_j(k) + \eta(k)\left(d_j(k) - \text{sign } w_j^T(k)\varphi(x(k))\right)\cdot\varphi(x(k)), \\ j &= 1,2,\ldots,m. \end{aligned} \tag{4}$$

Introducing further general criterion for all system outputs

$$E(k) = \sum_{j=1}^{m} E_j(k) = \sum_{j=1}^{m} e_j(k)u_j(k),$$

we can write learning algorithm of all system synaptic weights in form

$$W(k+1) = W(k) + \eta(k)\left(d(k) - \text{sign } W(k)\varphi(x(k))\right)\cdot\varphi^T(x(k)), \tag{5}$$

where $\text{sign}(u_1(k),u_2(k),\ldots,u_m(k))^T = (\text{sign}u_1(k),\text{sign}u_2(k),\ldots,\text{sign}u_m(k))^T$, $d(k) = (d_1(k),d_2(k),\ldots,d_m(k))^T$,

$W(k) = \begin{pmatrix} w_1^T(k) \\ w_2^T(k) \\ \vdots \\ w_m^T(k) \end{pmatrix} - (m\times h)$ is matrix of tuning synaptic weights.

It is known that gradient algorithms (3)-(5) provide the convergence in enough wide range of variation of learning rate parameter $\eta(k)$ [Derevitskiy, 1981], however at that convergence rate can be nonsufficient. Increasing of learning rate we can use quasi-Newton learning algorithms [Shepherd, 1997], for example,

$$w_j(k+1) = w_j(k) + \left(\varphi(x(k))\varphi^T(x(k)) + \eta I\right)^{-1} e_j(k)\varphi(x(k)), \tag{6}$$

where $\eta > 0$ is momentum term, $I - (\text{h}\times\text{h})$ is unity matrix.

Using lemma of matrix inversion we can show that [7]

$$\left(\varphi(x(k))\varphi^T(x(k)) + \eta I\right)^{-1}\varphi(x(k)) = \frac{\varphi(x(k))}{\eta + \|\varphi(x(k))\|^2},$$

and rewrite (6) in compact form

$$w_j(k+1) = w_j(k) + \frac{e_j(k)\varphi(x(k))}{\eta + \|\varphi(x(k))\|^2}, \tag{7}$$

or

$$W(k+1) = W(k) + \frac{d(k) - signW(k)\varphi(x(k))}{\eta + \|\varphi(x(k))\|^2}\varphi^T(x(k)), \tag{8}$$

for $\eta = 0$ this algorithm is multidimensional modification of optimal algorithm, introduced in [Tsypkin, 1984].

Conclusions

The diagnostic neuro-fuzzy system and its adaptive learning algorithm are introduced for solving pattern recognition, classification, diagnostics tasks etc under condition when training set value is comparable with input patterns dimension, and these patterns are fed for processing in on-line mode. The feature of proposed systems is significant smaller number of tuning parameters in comparison with the artificial neural networks, which solve the same task.

The system is characterized by simplicity of computational implementation, high speed of learning process, possibility of processing information, which is described in different scales (interval, ordinal, binary).

Bibliography

[Bodyanskiy, 2002] Ye. Bodyanskiy, Ye. Kucherenko, O. Chaplanov Diagnostic and prediction of time series using multilayer radial-basis neural network. Proc. 8 Russian conf. with internal. participation “Neurocomputers and its Applying”, Moskow, 2002, P. 209-213 (in Russian).

[Bodyanskiy, 2005] Ye. Bodyanskiy, Ye. Kucherenko, O. Mikhalev Petri Neuro-Fuzzy Networks in Modelling Tasks of Complex Systems, Dnepropetrivsk: Systemni Technologii, 2005, 311 p. (in Russian).

[Bodyanskiy, 2010] Ye. Bodyanskiy, O. Shubkina Semantic annotation of text documents based on hierarchical radial basis function network. Eastern-European Journal of Enterprise Technologies, 2010, 9(90), P. 70-74 (in Russian).

[Derevitskiy, 1981] D.P. Derevitskiy, A.L. Fradkov Applied Discrete Adaptive Control System Theory. M. Nauka, 1981, 216 p.

[Friedman, 2003] J. Friedman, T. Hastie, R. Tibshirani. The Elements of Statistical Learning. Data Mining, Inference and Prediction, Berlin: Springer, 2003, 552 p.

[Haykin, 1999] S. Haykin Neural Networks. A Comprehensive Foundation. Upper Saddle River, NJ: Prentice Hall, 1999, 842 p.

[Jang, 1997] J.-S.R. Jang, C.-T. Sun, E. Mizutani Neuro-Fuzzy and Soft Computing. Prentice Hall, Upper Saddle River, NJ, 1997, 640 p.

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

[Kurse, 2013] R. Kruse, C. Borgelt, F. Klawonn, C. Moewes, M. Steinbrecher, P. Held Computational Intelligence. A Methodological Introduction, Springer, 2013, 488 p.

[Ljung, 1999] L. Ljung System Identification: Theory for the User. PTR Prentice Hall, Upper Saddle River, N.J., 1999, 672 p.

[Osowski, 2006] S. Osowski Sieci neuronowe do przetwarzania informacji. Oficyna Wydawnicza PW, Warszawa, 2006.

[Shepherd, 1997] A.J. Shepherd Second-Order Methods for Neural Networks. London: Springer-Verlag, 1997, 145 p.

[Shynk, 1990] J.J. Shynk Performance surfaces of a single-layer perceptron. IEEE Trans. on Neural Networks, 1990, 1, P. 268-274.

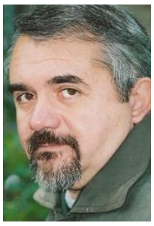
[Swamy, 2014] Ke-Lin Du, M.N.S. Swamy Neural Networks and Statistical Learning, Springer-Verlag London, 2014. - 824 p.

[Tsypkin, 1984] Ya.Z. Tsypkin Foundation of learning systems theory. M. Nauka, 1984, 320 p.

[Wang, 1992] L.X. Wang, J.M. Mendel Fuzzy basis functions, universal approximation, and orthogonal least squares learning. IEEE Trans. on Neural Network, 1992, 3, P. 807-814.

[Wang, 1994] L.-X. Wang Adaptive Fuzzy Systems and Control: Design and Stability Analysis. New Jersey: Prentice Hall, 1994, 256 p.

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 Electronic, Lenina av. 14, Kharkiv, Ukraine 61166, Tel +380577021890, bodya@kture.kharkov.ua

Vynokurova Olena - Doctor of Technical Sciences, Leading Researcher of the Control Systems Research Laboratory, Kharkiv National University of Radio Electronic, Lenina av. 14, Kharkiv, Ukraine, 61166, Tel. +380577021890, vinokurova@kture.kharkov.ua

Pliss Iryna - Candidate of Technical Sciences (Ph.D.), Leading Researcher of the Control Systems Research Laboratory, Kharkiv National University of Radio Electronic, Lenina av. 14, Kharkiv, Ukraine, 61166, Tel. +380577021890, pliss@kture.kharkov.ua

Peleshko Dmytro - Doctor of Technical Sciences, Professor of Department of Information Technology Publishing, National University “Lviv Polytechnic”, Stepan Bandera 28, buld. 5, Lviv, Ukraine, 79013, Tel. +38032258-27-79, dpeleshko@gmail.com