Behavioural nonlinear system models specified by various types of neural networks

E B Solovyeva

St. Petersburg Electrotechnical University, 5, Prof. Popov St., St. Petersburg, 197376, Russia

E-mail: selenab@hotbox.ru

Abstract. The nonlinear dynamic system modelling based on the input/output relationship results from solving the approximation problem. One can distinguish two large classes: polynomials and neural networks. The different types of neural networks draw attention. The advantages and disadvantages of various neural networks are emphasized. Represented analysis is important since it enables one to appreciate and choose a model, which is suitable for certain assigned conditions. The discrete time feedforward cellular neural network is suggested to filter non-Gaussian noise, as well as the example of nonlinear filters modelling to cancel the impulse noise is represented.

1. Introduction

The behavioural models of nonlinear dynamic systems are constructed using the sets of input and output signals as well as taking into account the information (knowledge) about these systems. Depending on the volume of a priori assigned knowledge, three levels are distinguished, namely, “white, grey and black box” [1, 2]. The less the volume of the information about a dynamic system is known, the darker the investigated object is getting. The “black box” level is established in the absence of any information about a system.

Let us highlight the “black box” approach. The reasons of turning to the “black box” methodology are the following [3]:

- the functional complexity of dynamic systems,
- the high level of the system integration,
- lack of the knowledge of the system parameters and characteristics,
- new high requirements to the parameters and characteristics of precision technique.

Some technical tasks can be formulated as problems, which are solved within the framework of the input/output relationship. Basing on the dynamic system description, the identification, modelling and synthesis can be characterized as follows.

The identification is the process of constructing a nonlinear operator (the determination of the model parameters by solving the optimization task) using the known set of input and output signals. The test input signals can be both deterministic and random ones.

The mathematical modelling is the using of the mathematical model, obtained in the identification process, for calculating nonlinear system output signals.

The synthesis is the creation of a hardware or software implemented device with a desirable goal, for example, for compensation, filtration, detection and so on [4, 5].
The stages of the “black box” modelling are the following:
– forming the sets of input and output signals of a modelled object,
– selecting the model type and its complexity,
– defining the model parameters on solving the approximation task with assigned error,
– validating a designed model.

The relationship between the input and output signal sets is described by the equation:

\[ y = F(y_{\text{past}}, u) \]  

Operator \( F \) in equation (1) approximates output \( y \) on the basis of previous output observation \( y_{\text{past}} \) and input \( u \).

According to the “black box” approach, the unique input/output relationship is established. The property of uniqueness means that there is no situation when any input signal from the corresponding set is mapped into two or more output signals [1–3].

2. Turning to neural models

The development of systems with very large-scale integration and widely practiced computer art caused immense interest in neural networks and substantial progress in their investigation in recent years.

On synthesizing neural networks, the complicated matter of the knowledge representation in the network is being solved. A priori information and invariants must be taken into account in order to simplify the architecture and learning process of the network. This problem plays the special part since the proper network configuration facilitates its specialization that is very important in view of the following reasons [6, 7]:
– the neural network having specialized structure usually comprises less number of free parameters, which are to be adjusted, than fully connected network does. Consequently, less amount of data is required for specialized network learning. Moreover, it takes less time, and such a network possesses much better generalization power;
– specialized networks are characterized by more network capacity;
– the implementation cost of specialized neural networks decreases because their size is significantly less than that of fully connected networks.

Nowadays, there is no clear solution of the problem how to construct specialized neural networks taking into account a priori information. Nevertheless, there are many forms of neural networks already developed, among which an acceptable structure (or structures) can be selected after estimation of their advantages and drawbacks and used for solving various research problems connected with maths, physics, engineering, neurobiology, psychology. In the course of developing the neural network theory, the sciences mentioned above will raise their theoretical and applied potentialities as well [6, 7].

3. Approximation properties of various neural networks

Let us consider the basic advantages and disadvantages of various types of neural networks used as the nonlinear models of dynamic systems.

3.1. Multi-layer perception network (MLPN)

The existence of hidden layers, which are not a part of the input or output of the network, enables one to solve complicated problems successively extracting the most important features (statistics of high orders) from the input signal. The ability of hidden neurons to elicit the statistical relation of high orders is especially essential when the input layer dimension is sufficiently large.

The mathematical description of the network is transparent, and thus it can be expressed as an algebraic dependence.

The disadvantages of the MLPN are the following:
• the high connectivity of the network (the accuracy approximation raising in one point of the network results in its impairment in another point of the network);
• the hidden neurons existence makes the network learning hard to visualization;
• the distributed nonlinearity form and the high network connectivity complicate significantly the theoretical analysis of the network.

3.2. Recurrent neural network (RNN)
The existence of feedbacks in the RNN leads to the weight coefficients decrease in comparison with multi-layer networks.

The disadvantage of the RNN is that, the existence of feedbacks results in the necessity of the network stability examination.

3.3. Radial-basis function network (RBFN)
Due to the local character of radial activation functions, it is easy to define the dependence between the basis functions parameters and the physical location of data for learning in multidimensional space. As a result, it is rather simple to obtain satisfactory initial conditions that is why the algorithm of the network learning is simplified.

Hybrid algorithm enables to separate the stage of the parameters selection of basis functions from the stage of obtaining the network weights, and hence it accelerates the network learning.

The disadvantages of the RBFN are as follows:
• the difficulty of the neuron quantity choice in a hidden layer. The small quantity of neurons does not allow one to decrease the approximation error to sufficient rate; on the other hand, the large quantity of them increases the computation complexity;
• on the network learning, “descents” can arise between individual Gaussian “bells”, whose branches overlap each other in some parts of the neural network model surface. If the distance between the “bells” centres is large and the ranges covered by branches are small, there can arise “dips” on the model surface caused by the local model nonsensivity to the input changes.

3.4. Wavelet neural network (Wavelet NN)
The orthonormalized wavelet-basis application increases the convergence rate of the network-learning algorithm.

The disadvantages of the Wavelet NN are the following:
• the high sensitivity of the convergence rate of the network-learning algorithm to initial conditions;
• the difficulty of the mother wavelet selection as an activation function among the large set of wavelet functions.

3.5. Functional link artificial neural network (FLANN)
The FLANN is a one-layer type that is why its learning algorithm comprises few transformations and ensures fast convergence to the approximation problem solution. The basis functions introduction aimed at lowering the condition number when the problem of approximation is being solved.

The disadvantage of the FLANN is that the approximation capabilities are less in a one-layer network than in multi-layer networks.

3.6. Polynomial perception network (PPN)
This network is a one-layer type so its algorithm of learning comprises few transformations and ensures fast convergence to the approximation problem solving.

The disadvantages of the PPN are as follows:
• the approximation capabilities are less in a one-layer network than in multi-layer networks;
• the use of the internal multidimensional converter of high degree results in the high error of the network learning.
3.7. **Spline neural network (Spline NN)**
The low degree nonlinearity in the Spline NN leads to the absence of numerical errors incident to the models of the high degree nonlinearity.

The disadvantage of the Spline NN is the large amount of the kept network parameters.

3.8. **Canonical piecewise-linear neural network (CPWLNN)**
The general (in canonical form) piecewise-linear model use leads to the reduction of the parameters number in comparison with separate linear simulation in space domains. In case of high nonlinearity, less amount of the network parameters is required in comparison with the Volterra polynomial.

The activation functions representation is suitable for implementation with a digital elementary basis (continuous sigmoid functions are replaced by the linear functions combinations).

The disadvantage of the CPWLNN is that, embedding absolute functions in one another generates the implicit form of functional model and, as a result, restricts the range of the nonlinear systems simulated by canonical network.

3.9. **Adaptive network based on fuzzy inference system (ANFIS)**
The compactness of the hierarchy fuzzy inference system, when there are few fuzzy rules in the hierarchy knowledge bases, enables one to describe adequately the multidimensional input/output relation.

The hybrid algorithm accelerates the network learning by separating the selection of the basis functions parameters from the obtainment of the network weights.

The disadvantage of the CPWLNN is the rise in the rules number and the membership functions parameters, when the number of the model inputs is increasing, results in swift growth in the approximation problem dimension.

4. **Example of nonlinear filters modelling to cancel non-Gaussian noise**
The problem of non-Gaussian noise filter synthesis is often effectively solved within the framework of the “black box” approach. According to this approach, the mathematical filter model describes the relationship between the sets of input and output signals. The model parameters are determined by solving the approximation problem in the mean-square norm.

Nonlinear filters are synthesized on the class of bit-map (dot element) half-tone images at the resolution measured by 256 grey levels, i.e., image is the matrix of integers (elements of brightness, pixels) in the interval [0; 255]. The pixel format is unit8. The impulse noise model is assigned as “salt and pepper” [8].

Four types of the nonlinear filter are applied to cancel the impulse noise in distorted images. Every filter is the cascade connection of the median filter with the 3x3 square aperture and a nonlinear unit in the form of either a neural network or polynomial.

The first type of the filter is suggested to build as the cascade connection of the median filter [8] and the discrete-time feedforward cellular neural network with the unity gain piecewise linear saturation function. Hereinafter, this filter is referred to as the combined discrete-time cellular neural network (CDTCNN).

The dynamics of a cell, \( C_{ij} \), in the feedforward cellular neural network is obtained by the following differential equation [9]:

\[
\frac{dx_{ij}(t)}{dt} = -x_{ij}(t) + \sum_{k=-1}^{1} \sum_{l=-1}^{1} b_{ij} u_{i+k,j+l}(t) + z
\]

where \( x_{ij}(t) \) denotes the state of cell \( C_{ij} \); \( u_{ij}(t) \) denotes the input of cell \( C_{ij} \) located in the sphere of influence with radius \( r \); \( t \) is the continuous time; \( b_{ij} \) is the feedforward synapse weight; \( z \) is the bias term.
The cell denoted by \( C_{ij} \) is located on position \((i, j)\) of the two-dimensional \( M \times N \) area, and its \( r \)-neighbourhood \( N'_{ij} \) is defined as follows:

\[
N'_{ij} = \{ C_{kl}, \max(|k-i|, |l-j|) \leq r, 1 \leq k \leq M; 1 \leq l \leq N \},
\]

where size \( r \) of the neighbourhood is a positive integer number.

Element \( u_{ij}(t) \) in equation (2) is obtained by moving a mask with the 3x3 window to position \((i, j)\) of the \( M \times N \) input image.

In the discrete time domain, equation (2) is turned into the equation:

\[
x_{ij}(n) = \sum_{k=-1}^{1} \sum_{l=-1}^{1} b_{kj} u_{i+k,j+l}(n) + z
\]

where \( n \) is the normalized discrete time.

Output signal \( y_{ij}(n) \) of cell \( C_{ij} \) results from treating by nonlinear activation function \( f(\cdot) \), which is usually specified as the unity gain piecewise linear saturation function described by the expression:

\[
y_{ij}(n) = f(x_{ij}(n)) = \frac{1}{2} \left( |x_{ij}(n)| + |x_{ij}(n)| - |x_{ij}(n)| \right).
\]

Eventually, the CDTCNN cell model is formed on the basis of equations (4) and (5) in the two-dimensional \( M \times N \) area described by equation (3).

The second type of the nonlinear filter synthesized is the cascade connection of the median filter [8] and the two-layer perceptron network with the hyperbolic tangent activation functions [6, 7], referred to as the combined two-layer perceptron network (CTLPN). The third type is the cascade connection of the median filter and the Volterra filter [1–3], referred to as the combined Volterra filter (CVF). The fourth type is the median filter (MF) performed at the 3x3 square aperture [8].

The parameters of the CDTCNN with five cells, the CTLPN including five neurons in the hidden layer and the CVF of the second degree are defined on solving the approximation problem in the mean-square norm while using the learning image with the size of 220x148 pixels. The length of learning sequence \( u(n) \) in equations (4) and (5) amounts to 32560 samples.

On filtering images, mean-square errors are estimated according to the formula:

\[
e = \frac{1}{Q} \sum_{n=1}^{Q} (y(n) - y^o(n))^2,
\]

where \( y(n) \) is the output signal of nonlinear filter, \( y^o(n) \) is a desirable signal, \( Q = 32 560 \).

Mean-square errors defined by equation (6) are summarized in table 1 under the impulse noise density equalled 0.5. “Tigers”, “Building” and “Fence” are the names of learning and two test images, correspondently. All the images have the size of 220x148 pixels. The synthesis of nonlinear filters was carried out on the basis of designing corresponding algorithms in MATLAB.

|        | CDTCNN | CTLPN | CVF  | MF   |
|--------|--------|-------|------|------|
| Tigers | 771    | 841   | 1206 | 2759 |
| Building | 1086   | 1186  | 1712 | 3014 |
| Fence  | 1558   | 1800  | 2214 | 3735 |

One can see from table 1 that the offered CDTCNN yields are higher filtration precision than the CTLPN, the CVF and the MF. It should be observed that the CDTCNN, the CTLPN and the CVF
provide different accuracy at the nearly equal complexity of these filters (56 parameters of the
CDTCNN and the CTLPN, 54 parameters of the CVF).

The use of the hyperbolic tangent activation functions in the CTLPN negatively affects the image
quality (white color turns to a gray one, as well as there is a bit ripple i.e. the image loses its
smoothness. Indeed, with an equal probability of the impulse noise (for instance, white and black dots
on images) occurrence, the filtration with different gains at low and high amplitudes of signals (in case
of the hyperbolic tangent) is not expedient.

In practice, the CDTCNN is more preferable in comparison with the CTLPN since its hardware
implementation in digital technique by means of signal processors or programmable logic devices is
simple due to using the piecewise linear saturation functions.

5. Conclusion
The mathematical modelling of nonlinear dynamic systems is frequently carried out basing on the
input/output system relationship by means of neural networks.

The represented analysis is useful for choosing the structure of the neural network a priori (prior to
its training or constructing the mathematical model of a nonlinear system). This aspect is important,
since the problem of the nonlinear system modelling is complicated in general formulation, therefore,
it is significant a priori to choose the model form, which is constructive for the mathematical
modelling under assigned conditions and provides the high approximation accuracy of the input/output
mapping of the nonlinear system.

The combined discrete time feedforward cellular neural network is proposed to filter the impulse
noise from half-tone images. This neural network with the piecewise linear saturation functions carries
out more accurate restoration of images in comparison with the combined two-layer perceptron
network comprising the hyperbolic tangent activation functions, the combined Volterra filter and the
median filter.

It should be emphasized that the hardware implementation of the cellular neural network is simpler
in comparison with the two-layer perceptron network since the piecewise linear saturation functions
used in the cellular network are simpler than the hyperbolic tangent activation functions included in
the perceptron network.

References
[1] Ogunfunmi T 2007 Adaptive Nonlinear System Identification, The Volterra and Wiener Model
Approaches (New York: Springer Science+Business Media, LLC) p 229
[2] Janczak A 2005 Identification of Nonlinear Systems Using Neural Networks and Polynomial
Models. A Block-oriented Approach (Berlin: Springer-Verlag Berlin Heidelberg) p 197
[3] Block-oriented Nonlinear System Identification 2010 ed F Giri and E-W Bai (Berlin: Springer-
Verlag Berlin Heidelberg) p 423
[4] Solovyeva E 2016 Proc. Int. Symp. on Problems of Redundancy in Information and Control
Systems (REDUNDANCY 2016) 150–154
[5] Solovyeva E B 2015 Radioengineering 24(4) 1071–1076
[6] Graupe D 2013 Principles of Artificial Neural Networks (New York: World Scientific
Publishing Co. Pte. Ltd.) p 363
[7] Haykin S 2009 Neural Networks and Learning Machines (New York: Pearson Education Inc.)
p 906
[8] Gonzalez R C and Woods R E 2008 Digital Image Processing (New York: Higher Education)
p 793
[9] Duan S, Hu X, Dong Z, Wang L and Mazumder P 2015 IEEE Trans. on Neural Networks and
Learning Systems 26(6) 1202–1213