Computer Applications in Electrical Engineering
Vol. 12 *2014*

A selected problem of the structure optimization and decomposition of the artificial neural network with cross-forward connections

Stanisław Płaczek
Akademia Finansów i Biznesu Vistula
00-725 Warszawa, ul. Chełmska 18A/28, e-mail: Stanislaw.placzek@wp.pl

The problem of an Artificial Neural Network (ANN) structure optimization is related to the definition of the optimal number of hidden layers and the distribution of neurons between layers depending on a selected optimization criterion and inflicted constrains. Using a hierarchical structure is an accepted default way of defining an ANN structure. The following article presents the resolution of the optimization problem. The function describing the number of subspaces is given, and the minimum number of layers, as well as the distribution of neurons between layers, shall be found. The structure can be described using different methods, mathematical tools, and software or/and technical implementation. The ANN decomposition into hidden and output layers – the first step to build a two-level learning algorithm for cross-forward connections structure – is described, too.

KEYWORDS: Artificial Neural Network, structure optimization, decomposition, coordination, cross connection

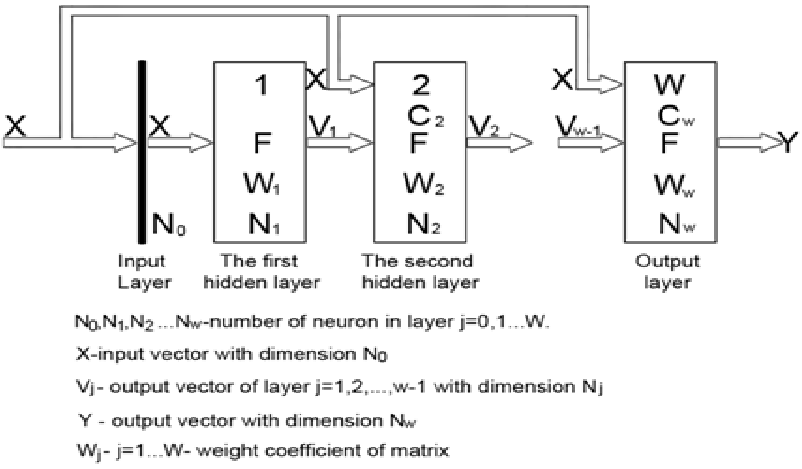
1. Network structure with cross connections

An Artificial Neural Network (ANN) is implemented as a universal approximator function with multidimensional variables. The selection of a neural network structure, aimed at the resolution of a specific problem is a challenging task. It is necessary to consider the following issues:

- The structure of a neural network, including the number of hidden layers and the distribution of neurons between layers. Usually, the size of an input and an output layer is defined by dimension of vectors X and Y respectively.
- The structure of the activation function, considering the requirements of a learning algorithm.
- The methods of data transfer between layers.
- The optimization criteria and type of a learning algorithm.

The most popular artificial neural network structure is the network with direct connection. This structure consist of at least one hidden layer. Data are fed from the proceeding layer to the succeeding one. The following paper consists of an analysis of the ANN with cross-forward connection (Fig. 1). In this structure, the input signal X is passed on to each layer in the network. Therefore, a layer j = 1,2,3….W, where W is the output layer, has two inputs:

- vector X, dimension N_0,
- vector V_{i-1} an output of the proceeding layer, dimension N_{i-1}.



$N_0, N_1, N_2 \ldots N_w$-number of neuron in layer j=0,1...W.

X-input vector with dimension N_0

V_j- output vector of layer j=1,2,...,w-1 with dimension N_j

Y - output vector with dimension N_w

W_j- j=1...W- weight coefficient of matrix

Fig. 1. A forward Connection ANN Structure

The signal transferred from an input layer to an output layer in the structure (Fig. 1), can be represented in the following equations:

$$U_i = W_i X \tag{1}$$

$$V_i = F(U) \tag{2}$$

$$E_i = W_i V_{i-1} + C_i X \tag{3}$$

$$Y = F(E_{Nw}) \tag{4}$$

where: $X[0:N_0]$ – an input vector, $W_i[1:N_1;1:N_0]$ – the matrix of weight coefficients of hidden layers, $U_i[1:N_i]$ – the analog signal of hidden layer, $V_i[1:N_i]$ – the output signal of hidden layer, $C_i[1:N_1;1:N_0]$ – the matrix of weight coefficients of Cross Connections, $Y[1:N_w]$ – the output signal of output layer, i = 1,2,…,w-1 – the number of hidden layers.

The initial feature space X of dimensionality N_0 is divided by every neuron in hidden and output layers into N_i subspaces. In accordance [2, 8], the total number ψ of subspaces for an ANN with "W" hidden and output layers is calculated:

$$\psi(N_0, W) = \prod_{i=1}^{W} \phi(N_0, N_i) \tag{5}$$

$$\phi(N_0, N_i) = C_{N_i-1}^{N_0} + 2\sum_{k+0}^{N_0-1} C_{N_i-1}^{k} \tag{6}$$

$$C_n^k = \frac{n!}{k!(n-k)!} \qquad \text{and} \qquad C_n^k = 0 \qquad \text{when } k > n \qquad (7)$$

Table. 1. Number of subspace $\phi(N_o,N_i)$

N_i $/N_0$	1	2	3	4	5	6	7	8	9	10
1	2	2	2	2	2	2	2	2	2	2
2	3	4	4	4	4	4	4	4	4	4
3	4	7	8	8	8	8	8	8	8	8
4	5	11	15	16	16	16	16	16	16	16
5	6	16	26	31	32	32	32	32	32	32
6	7	22	42	57	63	64	64	64	64	64
7	8	29	64	99	120	127	128	128	128	128
8	9	37	93	163	219	247	255	256	256	256
9	10	46	130	256	382	466	502	511	512	512
10	11	56	176	386	638	848	968	1013	1023	1024

The number of subspaces formed by the division of N_0 dimensional input vector X by N_i neurons present in the hidden or the output layer is shown in Table 1, which can be divided into two parts:

- the upper right corner above the diagonal accomplished the relation $N_i \le N_0$ – the dimensionality of an input vector X is greater than/equal to the number of neurons in each hidden layer N_i,
- the left down corner below the diagonal accomplished the relation $N_i > N_0$ – the number of neurons in each hidden layer N_i is greater than the number of neurons in the input space X.

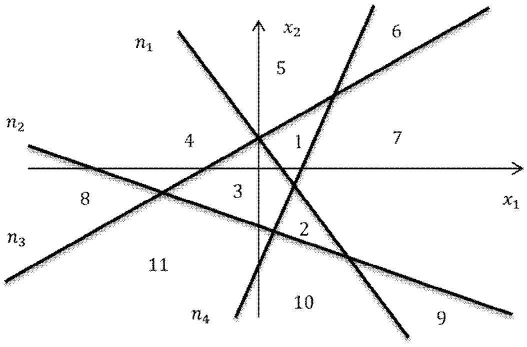


Fig. 2. Two dimensional input space N_0 is divided into 11 subspaces by $N_i = 4$ neurons in one hidden layer

Example 1. For ANN (2-4-2) which includes one hidden layer with $N_1 = 4$ neurons and an output layer with $N_w = 2$ neurons. The total number of subspaces can be calculated using formula (5) and Table 1.

$$\psi(2,2) = \prod_{i=1}^{W} \phi(2, N_i) = \phi(2,4)\phi(2,2) = 11 \cdot 4 = 44$$

Taking into account formulas (5), (6) we can formulate the next optimization tasks:

- The total number of neurons N in ANN is given

$$N = \sum_{i=1}^{w-1} N_i$$

We should find the maximum number of subspaces ψ, the optimal number of the hidden layer (W-1) and the distribution of neurons N_i between layers,

- The number of subspaces ψ is given. We should find the number of hidden layers (W-1), the distribution of neurons between layers N_i and the total number of neurons N.

2. The issue of structure optimization

Total number of ψ subspaces for cross connection network is given. The minimum number of (W-1) layers and N_i neuron distribution between layers shall be found. The target function is defined:

$$\min N = \min_{w-1, N_i} \sum_{i=1}^{w-1} N_i \tag{8}$$

and constrains

$$\psi = \prod_{i=1}^{w-1} \phi_i(N_0, N_i) \tag{9}$$

$$\phi(N_0, N_i) = C_{N_i-1}^{N_0} + 2 \sum_{k+0}^{N_0-1} C_{N_i-1}^{k} \tag{10}$$

Additionally we assumed, that

$$N_i > N_0 \qquad \text{for all} \quad i = 1,2,\ldots,W\text{-}1 \text{ layers} \tag{11}$$

In default way it is assumed that dimensionality of both input N_0 and output Nw vectors of ANN are given. Our objective is to minimize the total number of neurons and layers. This complex problem could be solved regarding the relation between dimensionality of feature space, N_0, and number of neurons in each of N_i hidden layers.

For Kuhn – Tucker condition imply that the following constraints (11) could be written:

$$N_i \geq N_0 + 1 \tag{12}$$

To find solution, Kuhn – Target condition could be applied. Taking into account (8), (9), (12) Lagrange equation is written

$$L = \sum_{i=1}^{H} N_i + \lambda_0 \left[\psi - \prod_{i=1}^{H} \phi_i(N_0, N_i) \right] + \sum_{i=1}^{H} \lambda_i (N_i - N_0 - 1) \qquad (13)$$

where: H = W - 1. Set of equations could be written

$$\frac{\partial L}{\partial N_i} = 1 - \lambda_0 \frac{\partial \phi}{\partial N_i} \prod_{i=1}^{H} \phi(N_0, N_i) + \lambda_i = 0 \qquad \text{for i = 1,2,...H} \qquad (14)$$

$$\frac{\partial L}{\partial \lambda_0} = \psi - \prod_{i=1}^{H} \phi(N_0, N_i) = 0 \qquad (15)$$

$$\frac{\partial L}{\partial \lambda_i} = N_i, - N_0 - 1 = 0 \qquad \text{for i = 1,2,...H} \qquad (16)$$

From (16) N_i could be found

$$N_i = N_0 + 1 \qquad \text{for i = 1,2,...H} \qquad (17)$$

Using (14) and (15) formula (14) could be rewritten

$$\frac{\partial L}{\partial N_i} = 1 - \lambda_0 \frac{\partial \phi}{\partial N_i} \psi + \lambda_i = 0 \qquad \text{for i = 1,2,...H} \qquad (18)$$

and finally

$$\lambda_i = \lambda_0 \frac{\partial \phi}{\partial N_i} \psi - 1 \qquad \text{for i = 1,2,...H} \qquad (19)$$

and finally

$$H = \frac{\ln \psi}{\ln[\phi(N_0, N_0 + 1)]} \qquad (20)$$

Using (18) and (8) the minimum sum of neurons distribution

$$N = H(N_0 + 1) \qquad (21)$$

The aforementioned means that

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_H \qquad (22)$$

Using (18) and (15) formula (15) could be rewritten

$$\psi = \phi(N_0, N_0 + 1)^H \qquad (23)$$

Example 2. Number of subspaces ψ= 100 and feature space dimensionality is N_0 = 3. Find number of hidden layers H and total number of neurons accomplishing the number of ψ subspaces. We assumed, that for all hidden layers.

From (17) $\qquad N_i = N_0 + 1 = 3 + 1 = 4$

From (22) calculate

$$H = \frac{\ln \psi}{\ln[\phi(N_0, N_0 + 1)]} = \frac{\ln 100}{\ln[\phi(3,4)]} = \frac{\ln 100}{\ln 15} = 1{,}7 \approx 2$$

From (23) calculate total number of neuron $N = 2 \cdot 4 = 8$.

The first step is the optimal structure of the ANN with cross connection. To describe the structure, independently from the ANN complexity, and in a default way, the partition on layers is used: the input layer, one or more hidden layers, and the output layer. The input layer and the output layer connects the ANN with the external world (the environment) and is usually done by the solving task. Next step, it is connected with the learning algorithm structure.

3. Structure decomposition of the ANN with cross connections

The learning algorithm acceleration can be achieved by dividing the more complicated algorithm into smaller parts and using parallel procedures In the next step one can use, for example, the modern programming tools or thread programming. From the algorithm point of view, a learning algorithm is multiplying matrices of weight coefficients for forward calculations. For back calculation, matrices of differential coefficients are multiplying, too. These processes could be time-consuming, especially when the dimensionality of the learning task is quite substantial. As an example, one may take into account the cross connection ANN with one hidden layer.

Preliminary analyses of the ANN structure shown on the existing hierarchical structure (Fig. 3).

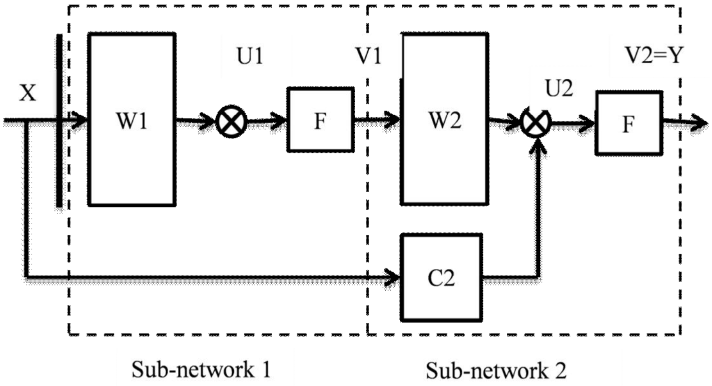


Fig. 3. The decomposition of ANN with one hidden layer and output layer

One can locate the first sub-network between the input vector X and the internal vector V1. It is defined by one matrix W1. The vector X, dimensionality N_0 can be preliminary filtering or normalizing in the input layer. According to the literature describing the ANN structure, the input layer does not include the total number of layers. Only all hidden layers and the output layer are built in. The second sub-network is located between the vector V1 and the output vector V2 or Y. It is defined by two matrices, W2 and C2. The decomposition of the ANN with cross connection relies on separating the sub-network one and sub-network two. This can be done by introducing two sets of vectors into the ANN structure:

– V1 and V12: V1 is calculated by the first sub-network and it is the first parameter of the local target function Φ1. V1 is sending to the coordinator which is using its own algorithm to calculate vector V21 and is sending it to the second sub-network.
– V2 and V21: V21 is calculated by the coordinator. The coordinator is using its own algorithm and parameters, and sending by the second sub-network can have the ability to calculate the second parameter of the local target function Φ1. vector V2 can be temporarily variable.

To summarize, one can build the new algorithm structure using a hierarchical scheme (Fig. 4).

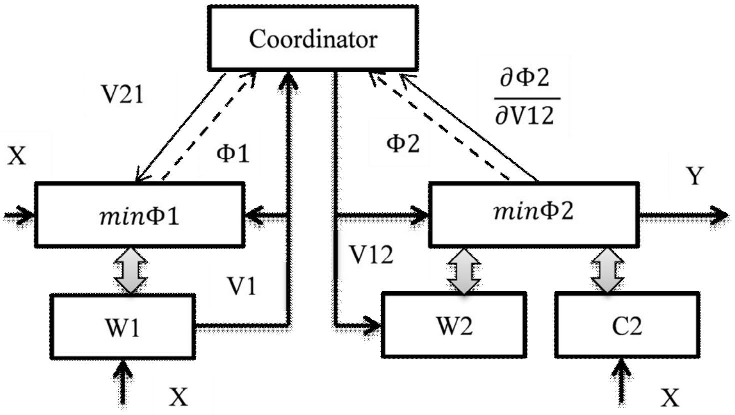


Fig. 4. Coordination scheme

In Fig. 4 one can see two levels. On the first level, two sub-networks are seen, representing the hidden and the output layer respectively. One needs to bear in mind that the term “level” is used to describe the hierarchical structure of the ANN. The term “layer”, as primary, is used to describe the ANN structure in the standard sense.

Additionally, in order to analyse the first level, the sub-network one with the local target function Φ1 can be divided into N_1 independent tasks. In the same way, the second sub-network with the local target function Φ2 can be divided into N_2 independent tasks. These can be used to build a parallel algorithm.

One should note that two kinds of vertical interaction between the first and second level are used [10]. One is downward from the second level to the first; optimization parameters for both local target functions Φ1, Φ2 respectively. The other kind of vertical interaction is upward. The information is sent from the first level to the second. It is a feedback signal.

The coordinator can use its own algorithm to achieve the global target function. One should stress that in a two-level learning algorithm, the three target functions are defined.

– The global target function for all structures and tasks

$$\psi(W2,V2)=\sum_{k=1}^{k+N_2}\psi_k=\sum_{k=1}^{k=N_2}(y_k-z_k)^2 \tag{24}$$

– The local target functions for the first level

$$\phi 1=\sum_{i=1}^{i=N_2}\phi 1_i=\sum_{i=1}^{i=N_2}(\nu 1_i-\nu 21_i)^2 \tag{25}$$

$$\phi 2=\sum_{i=1}^{i=N_2}\phi 2_i=\sum_{i=1}^{i=N_2}(\nu 12_i-z_k)^2 \tag{26}$$

– The coordinator target function. To connect sub-networks on the first level, two functions are used.

$$V12=G(V1) \tag{27}$$

$$V21=H(V2) \tag{28}$$

4. A learning algorithm of the ANN with cross connections

4.1. The first level tasks

For the first level, the minimum of the two local target functions is seached.

$$\min \phi 1=\frac{1}{2}\sum_{i=1}^{N_2}\left(f\left[\sum_{j=0}^{j=N_0}W1_{ij}*x_j\right]-\nu 21_i\right)^2 \tag{29}$$

$$\frac{\partial\phi 1}{\partial w1_{ij}}=(\nu 1_1-\nu 2_i)*\frac{\partial f}{\partial e_k}*x_j \tag{30}$$

$$e_i=\sum_{j=0}^{j=N_0}W1_{ij}*x_j \tag{31}$$

$$\nu 1_i=\frac{1}{1+\exp^{-\alpha+e_i}} \tag{32}$$

$$w1_{ij}(n+1) = w1_{ij}(n) - \alpha_1 * \frac{\partial \phi 1_i}{\partial W1_{ij}} \tag{33}$$

where $\nu 21_i$ is a parameter calculated by the coordinator.

$$\min \phi 2 = \frac{1}{2}\sum_{k=1}^{N_2} \left(f\left[\sum_{i=0}^{N_0} W2_{ki} * \nu 12_i + \sum_{j=0}^{N_0} C2_{kj} * x_j \right] - z_{ki} \right)^2 \tag{34}$$

$$\frac{\partial \phi 2}{\partial w2_{ki}} = (\nu 2_k - z_k) * \frac{\partial \nu 2_k}{\partial u2_k} * \nu 12_i \tag{35}$$

$$\frac{\partial \phi 2}{\partial w2_{kj}} = (\nu 2_k - z_k) * \frac{\partial \nu 2_k}{\partial u3_k} * x_j \tag{36}$$

$$\frac{\partial \phi 2}{\partial w21_i} = \sum_{k=1}^{N_2} (\nu 2_k - z_k) * \frac{\partial \nu 2_k}{\partial u2_k} * W2_{ki} \tag{37}$$

$$W2_{ki}(n+1) = W2_{kij}(n) - \alpha_2 \frac{\partial \phi 2}{\partial W2_{ki}} \tag{38}$$

$$C2_{kj}(n+1) = C2_{kj}(n) - \alpha_2 \frac{\partial \phi 2}{\partial W2_{kj}} \tag{39}$$

where

$$u2_k = \sum_{i=0}^{N_1} W2_{ki} * \nu 12_i \tag{40}$$

$$u3_k = \sum_{j=0}^{N_0} C2_{kj} * x_j \tag{41}$$

The first level tasks should achieve their minimum value of the local target functions Φ1, Φ2 depending on the coordination parameters V21 and V12. These parameters are calculated by the coordinator in every iteration.

4.2. The second level task – coordinator task

For the coordinator, two functions G and H are defined which transform signals V1 → V12 and V2→ V21. At the same time, the coordinator should have the ability to change the value of learning coefficients α_1, α_2 and γ_2 by using transformation functions $h_1 1(\Phi 1,\Phi 2)$ and $h_1 2(\Phi 1,\Phi 2)$.

$$\nu 12_i = \lambda_1 \nu 1_i \tag{42}$$

$$\nu 2_i(n+1) = \nu 12_i(n) - \gamma_2 \frac{\partial \phi 2}{\partial \nu 12_i} \tag{43}$$

$$\nu 21_i = \lambda_2 2 \tag{44}$$

$$\alpha_1 1(n+1) = \alpha_1 1(n) + h_1 1(\phi 1, \phi 2) \tag{45}$$

$$\gamma_1 2(n+1) = \gamma_1 2(n) + h_1 2(\phi 1, \phi 2) \tag{46}$$

5. The example

The main dynamic characteristics of the learning processes are shown in the following example. The stress is laid on the first-level task: finding the minimum of the local target functions Φ1,Φ2. The structure of the ANN with cross connections is simple and can be described as the ANN (3-5-1).

This means that the ANN includes 3 input neurons, 5 neurons in one hidden layer, and 1 output neuron. The first exists when the learning process achieves 1900 iteration number. The second maximum is achieved when the learning process is close to 5500 iteration number.

The sigmoid activation functions are implemented in both the hidden and the output layers. Three arguments of the XOR function are fed as the input data (Fig. 5), which shows how the target function of the second sub-network changes its value during the learning process. This characteristic correlates with the process (Fig. 6), but the quality of the characteristic is different. In Fig. 6 one can observe two local maximums.

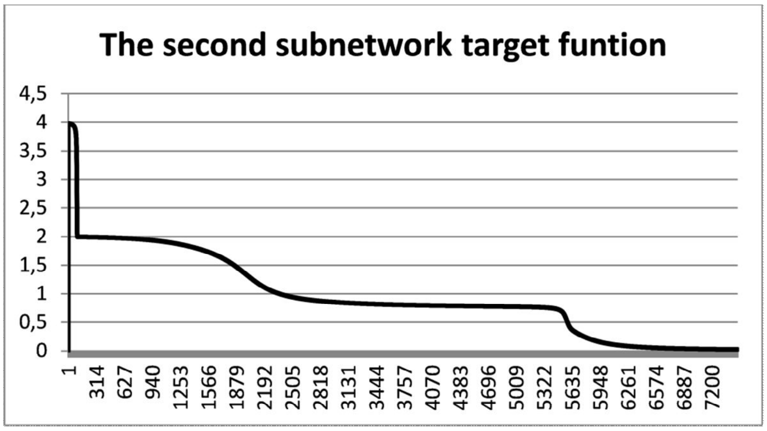


Fig. 5. The dynamic characteristic of the second target function Φ2

At these two points, the characteristic of the second local target function decreases its value quit dramatically and finally, in an asymptotic way, achieves the minimum value. This can be explained thusly: in the first stage of the learning process the matrix W1 has to stabilize its coefficients and the local target function Φ1 increases its value to accelerate this process. Afterwards, the stable state is achieved and the second sub-network stables its matrix W2 coefficients as well. The local target function Φ2 achieves its minimum value.

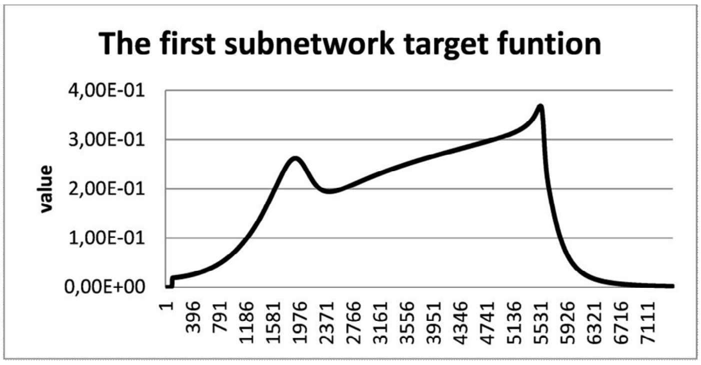


Fig. 6. The dynamic characteristic of the first target function Φ1

6. Conclusion

The structure of an ANN with cross connection enables optimizing its structure using different optimizing criteria. When data from the input vector X are send to the output layer, the dynamic characteristics of the learning process are changed. The second layer, using extra information, stabilizes its matrix coefficient faster. With decomposition and a simple coordination algorithm one can use a quite simple learning algorithm which can also work faster. All in all, an ANN with cross connection has very interesting characteristics and should be studied in future work.

References

[1] S. Osowski, Sieci neuronowe do przetwarzania informacji, Oficyna Wydawnicza Politechniki Warszawskiej, Warszaw 2006.

[2] O. B. Lapunow, On possibility of circuit synthesis of diverse elements, Mathematical Institut of B.A. Steklova, 1958.

[3] Toshinori Munakate, Foundational of the New Artificial Intelligence, Second Edition, Springer 2008.

[4] Colion Fyle, Artificial Neural Network and Information Theory, Department of Computing and Information System, The university of Paisley, 2000.

[5] Joarder Kamruzzaman, Rezaul Begg, Artificial Neural Network in Finance and Manufacturing, Idea Group Publishing, 2006.

[6] L. Rutkowski, metody i techniki sztucznej inteligencji, Wydawnictwo naukowe PWN, Warszawa 2006.

[7] S. Placzek, B. Adhikari, Analysis of Multilayer Neural Network with Direct Connection Cross-forward Connection, Conference CS&P 2013 Warsaw University, Warszawa 2013.

[8] W. Findeisen, J. Szymanowski, A. Wierzbicki, Teoria i metody obliczeniowe optymalizacji. Państwowe Wydawnictwo Naukowe, Warszawa 1977.

[9] Ch. M. Bishop, Pattern Recognition and Machine Learning, Springer Science + Business Media, LLC 2006.

[10] M. D. Mesarocic, D. Macko, and Y. Takahara, Theory of hierarchical multilevel systems, Academic Press, New York and London, 1970.

[11] Zeng-Guang Hou.Madan M.Gupta, Peter N. Nikiforuk, Min Tan, and Long Cheng, A Recurrent Neural Network for Hierarchical Control of Interconnected Dynamic Systems, IEEE Transactions on Neural Networks, Vol. 18, o. 2, March 2007.