Computer Applications in Electrical Engineering
Vol. 13 *2015*

"P" coordinator scheme and interaction prediction principle in hierarchical structure of ANN

Stanisław Płaczek
Vistula University
02-787 Warszawa, ul. Stokłosy 3, e-mail: Stanislaw.placzek@wp.pl

When implementing the hierarchical structure [4][5] of the learning algorithm of an Artificial Neural Network (ANN), two very important questions have to be solved. The first one is connected with the selection of the broad coordination principle. In [1], three different principles are described. They vary with regard to the degree of freedom for the first-level tasks. The second problem is connected with the coordinator structure or, in other words, the coordination algorithm. In the regulation theory, the process of finding the coordinator structure is known as the feedback principle. The simplest regulator structure (scheme) is known as the proportional regulator – "P" regulator. In the article, the regulator structure and its parameters are analysed as well as their impact on the learning process quality.

KEYWORDS: Artificial Neural Network (ANN), hierarchy, decomposition, coordination, coordination principle, P-regulator, feedback principle

1. Calculation complexity

For a multi-layered ANN a lot of hidden layers and an output layer can be sectioned off [3]. Every layer has its own output vector that is an input vector of the next layer. For a standard two-layer network, both the hidden layer and the output layer can be described as sub-networks. These sub-networks form the first level of the hierarchical structure. Now the network consists of the two of sub-networks, for each of which a local target function is defined by $\phi = (\phi_1, \phi_2)$. Similarly to the ANN structure decomposition, a learning algorithm using error backpropagation can be also decomposed. We can sort out:

- The first-level task in which the minimum of the local target functions ϕ_1, ϕ_2 is searched.
- The second level task, which has to coordinate all the first-level tasks.

In such a learning algorithm there is a set of optimization tasks on the first level.

These tasks are searching for the minimum value of the target function ϕ_1, ϕ_2. Unfortunately, these are non-linear tasks without constraints. In practice, standard procedures to solve these problems exist. But in the two-level learning algorithm structure, the coordinator is not responsible for solving the global task ϕ. As noted

above, the coordination principles were developed to solve this complicated problem.

In [2], the most popular interaction prediction principle was implemented. In this principle, the coordinator plays an active role in the current ANN learning process. In each iteration the coordinator and all of the first-level sub-tasks interchanged information. The first-level sub-tasks are the optimal searching tasks. Usually they look for the minimum of the Mean Squared Error (MSE). The coordination algorithm structure is primary related to the interaction prediction principle and two signals are used: primary from up to down $\gamma = (\gamma_1, \gamma_2)$ and the feedback signal from down to up $\varepsilon = (\varepsilon_1, \varepsilon_2)$. The primary signals are known as coordination signals and are sent from the coordinator to all of the first-level sub-tasks. In this way the coordinator helps them to solve their own optimization tasks. The forecasting (predicted) coordination signals could be not exactly correct and the sub-tasks calculate their own coordination signals value. These signals are sent from down to up into the coordinator. The coordinator uses its own gradient method to calculate the new value of coordination signals (improving its own previous decision). The process could be continued as the coordinator and all the first-level sub-tasks have solved their own tasks (find minimum errors value) (Fig. 1).

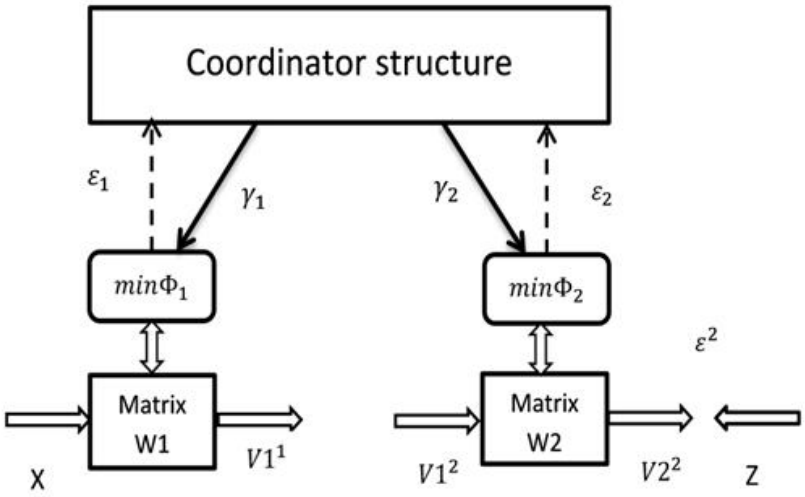

Fig. 1. Decomposition and coordination scheme

In the hierarchical learning algorithm, the next target functions can be defined as:

- The global target function ϕ,
- Two local target functions ϕ_1, ϕ_2,
- The coordinator target function ψ.

2. Two–level learning algorithm structure

In the decomposed ANN structure [6] one can define the next target functions:

- Global target function. For all epoch:

$$\phi(W1,W2,X,Y)=\frac{1}{2}\sum_{k=1}^{N_2}\sum_{p=1}^{N_p}(y_k^p-z_k^p)^2 \tag{1}$$

where: $Y[1: N_2, 1: N_p]$ – the ANN output value, $Z[1: N_2, 1: N_p]$ – the teaching data, N_2 – the number of output neurons, N_p – the dimensionality of the training set.

- Local target function Φ_1. For all epoch:

$$\phi_1(W1,X,\gamma)=\frac{1}{2}\sum_{p=1}^{N_p}\sum_{i=1}^{N_1}(v1_i^p-\gamma_i^p)^2=\frac{1}{2}\sum_{p=1}^{N_p}\sum_{i=1}^{N_1}(f(\sum_{j=0}^{N_0}W1_{ij}*x_j)-\gamma_i^p)^2 \tag{2}$$

where: $\gamma[1: N_1, 1: N_p]$ – the coordination matrix as parameters, N_1 – the number of hidden neurons, N_0 – the number of input neurons.

- The local target function Φ_2. For all epoch:

$$\phi_2(W2,Z,\gamma)=\frac{1}{2}\sum_{p=1}^{N_p}\sum_{k=1}^{N_2}(v2_k^p-z_k^p)^2=\frac{1}{2}\sum_{p=1}^{N_p}\sum_{k=1}^{N_2}(f(\sum_{i=0}^{N_1}W2_{ki}*\gamma_k^p)-z_k^p)^2 \tag{3}$$

where: $\gamma[1: N_1, 1: N_p]$ – the coordination matrix as an input variable, N_2 – the number of output neurons. f(∗) – a sigmoid function.

Using (2), one can calculate the feedback signal $\varepsilon 1_i^p$ and the new value of matrix W1.

$$\varepsilon 1_i^p=f(\sum_{j=0}^{N_0}W1_{ij}*x_j) \tag{4}$$

$$\frac{\partial\phi_1}{\partial W1_{ij}}=\sum_{p=1}^{N_p}(v1_i^p-\gamma_i^p)*f'*x_j^p \tag{5}$$

$$W1_{ij}(n+1)=W1_{ij}(n)-\alpha_1*\frac{\partial\phi_1}{\partial W1_{ij}} \tag{6}$$

For the second sub-network using (3), one can calculate the new value of $\varepsilon 2_i^p$ and the new value of matrix $W2_{ki}$.

$$\frac{\partial \phi_2}{\partial W2_{ki}} = \sum_{p=1}^{N_p} (v2_k^p - z_k^p) * f' * \gamma_k^p \tag{7}$$

$$W2_{ki}(n+1) = W2_{ki}(n) - \alpha_2 * \frac{\partial \phi_2}{\partial W2_{ki}} \tag{8}$$

$$\frac{\partial \phi_2}{\partial \gamma_i^p} = \sum_{k=1}^{N_1} (v2_k^p - z_k^p) * f_2' * W2_{ki} \tag{9}$$

$$\varepsilon 2_i^p(n+1) = \gamma 2_i^p(n) - \alpha_3 * \frac{\partial \phi_2}{\partial \gamma 2_i^p} \tag{10}$$

where: α_3 – the learning coefficient.

To summarize the calculations above we can conclude that the first-level, obviously, includes two sub-networks and two optimizations tasks. The first sub-network calculates the new matrix coefficient $W1_{ik}(n+1)$ and the feedback signal $\varepsilon 1_i^p$ value by taking from the coordinator the parameter γ_k^p and using the optimization procedure. The feedback signal is sent into the coordinator. For the second sub-network, the coordination signal γ_k^p sets the optimization procedure in motion and calculates the new matrix coefficient $W2_{ki}(n+1)$ and the feedback signal $\varepsilon 2_k^p$ value. The feedback signal is sent into the coordinator, too. After that, the coordinator procedure has to calculate the new coordinator signal $\gamma_k^p(n+1)$. Thus, the coordinator structure is the main problem in a two-level learning algorithm.

3. The coordinator structure

According to Fig. 1, the coordinator supervised two discrete dynamic objects: two first-level sub-networks. In a discrete moment of time every object has the input data calculating the output value and the appropriate derivative value. After that they are ready to modify the weight coefficients matrices W1, W2.

At the same discrete moment of time, objects calculate the feedback signals values that are sent to the coordinator. The conclusion is simple – the coordinator is a discrete dynamic object, too [7]. It contains two input signals and one output (Fig. 2).

All parts of the ANN learning processes are dynamic and, as we can infer, non-linear. Non-linearity is connected with all the activation functions that could be unipolar or bipolar. The learning process is an iteration process and during the discrete learning time, the coordinator parameters could be far away from their optimal theoretic value. The main target of the adaptation level is achieving data from the coordinator and using its own statistic procedure to

change the coordinator parameters. This process could be time-consuming and should be used for complicated coordinator structures.

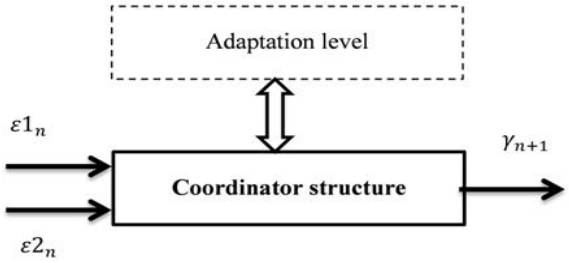

Fig. 2. Simplified coordinator structure

An input–output coordinator description can be achieved by a discrete differential equation:

$$\gamma_k(n+1) = \varphi[\varepsilon 1_i(n), \varepsilon 2_i(n)] \tag{11}$$

where: $\varepsilon 1_i(n)$, $\varepsilon 2_i(n)$ – the coordinator input, $\gamma_k(n+1)$ – the coordinator output.

For a linear coordinator structure, the equation (11) has a standard linear discrete structure.

In the article, the simplest coordinator structure is investigated: the proportional regulator [7], which is known as "P" regulator, with the internal feedback (Fig. 3).

Input signals $\varepsilon 1_i(n)$, $\varepsilon 2_i(n)$ are handed to differential blocks and are compared with the coordinator signal $\gamma_k(n)$. This signal is delayed into one moment in time using two delayed blocks.

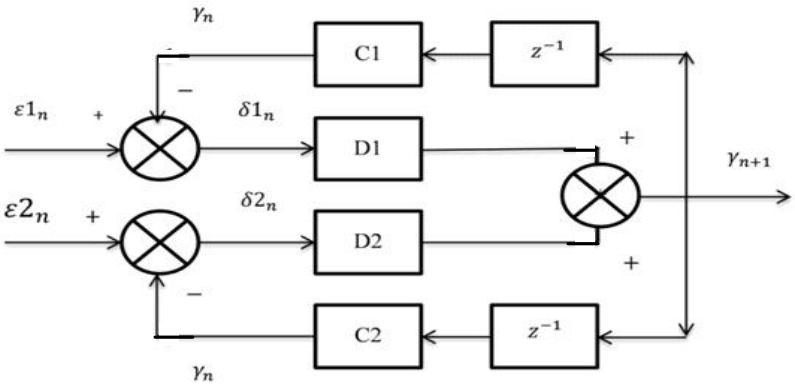

Fig. 3. Full coordinator structure

Next, two signals $\delta 1_i(n)$, $\delta 2_i(n)$ are handed into forward blocks to be amplified into the correct value. The final signal is achieved by a summarized block.

Using Figure 3, the subsequent set of equations could be written:

$$\delta 1(n) = \varepsilon 1(n) - C1 \cdot \gamma(n) \tag{12}$$

$$\delta 2(n) = \varepsilon 2(n) - C2 \cdot \gamma(n) \tag{13}$$

$$\gamma(n+1) = D1 \cdot [\varepsilon 1(n) - C1 \cdot \gamma(n)] + D2 \cdot [\varepsilon 2(n) - C2 \cdot \gamma(n)] \tag{14}$$

After simple math operations, the final formula has the following structure:

$$\gamma(n+1) = [D1 \cdot \varepsilon 1(n) + D2 \cdot \varepsilon 2(n)] - \gamma(n) \cdot [C1 \cdot D1 + C2 \cdot D2] \tag{15}$$

The coordinator has two channels with the local feedback. To use this formula four different parameter values should be defined. It is not a simple task. To simplify this, the next assumption is used:

$$C1 = C2 = C \tag{16}$$

$$D1 = D2 = D \tag{17}$$

The regulator structure is symmetrical with the same feedback and forward parameters.

Ultimately, the final formula is achieved:

$$\gamma(n+1) = D[\varepsilon 1(n) + \varepsilon 2(n)] - 2 \cdot C \cdot D \cdot \gamma(n) \tag{18}$$

These two "C" and "D" parameters are correlated; neither could be too big or too small. The task of finding the optimal regulator parameters value is sent into the adaptation level, which will be studied in the future.

4. Numerical example

In the following example the main dynamic characteristics of the learning process are shown. The emphasis is put on the characteristics of the first-level local target functions, ϕ_1, ϕ_2, and the second-level: the coordinator target function ψ. The structure of the ANN is simple and can be described as ANN (3-5-1). This means that the ANN includes 3 input neurons, 5 neurons in the hidden layer and 1 output neuron. Sigmoid activation functions are implemented in both hidden and output layers. Three arguments of XOR3 function are fed as input data. So, every epoch includes 8 vectors.

Figure 4 shows the output characteristics of two sub-networks. The target function ϕ_2 decreases the error value faster than ϕ_1.

It can be said that the hidden layer (the target function ϕ_1) consists of more neurons (5 neurons) than the output layer (only 1 neuron). The quality of the output characteristics depends on the coordinator parameters – feedback "C" and forward "D".

To measure the coordinator characteristics the subsequent formula is used:

$$\psi = \sum_{p=1}^{N_p} [(\gamma^p - \varepsilon_1^p)^2 + (\gamma^p - \varepsilon_2^p)^2] \tag{19}$$

Figure 5 shows two main characteristics: the global target function for all ANNs and the coordinator target function. Comparing Figures 4 and 5, one can notice that only the global target function ϕ has different characteristics. At the beginning their value increases over "2" and after that slowly decreases to zero. At about 2000 iterations numbers the ANN learning process achieves its target value.

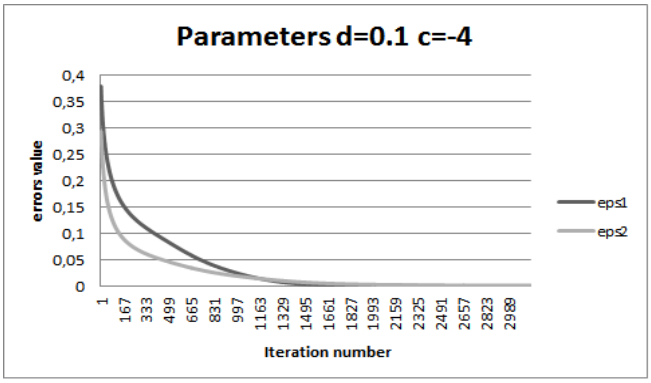

Fig. 4. Value of the local target functions ϕ_1, ϕ_2, depending on the iteration number
eps1- the first subnetwork target function value
eps2- the second subnetwork target function value

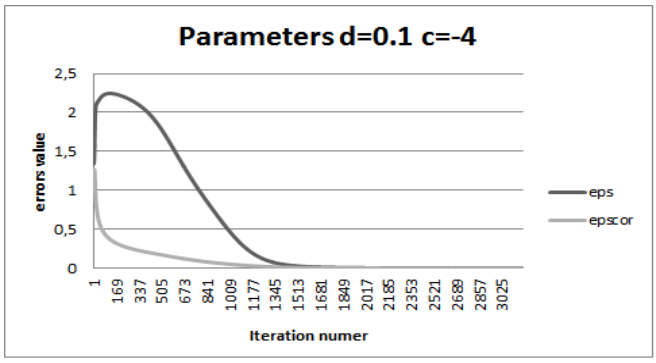

Fig. 5. Value of the global target function ϕ and coordinator target function ψ, depending on the iteration number
eps- all network target function value
epscor – coordinator target function value

The coordinator plays a very important role in the learning process, it could be even said that the principal one. Two input values ε_1^p and ε_2^p are different from the coordinator signal value γ^p. The differences are substantial at the beginning of the learning process (at the beginning of the iteration process). In Fig. 6, the coordinator signal γ^p increases its own value. The feedback signal ε_1^p

has to increase its value as well. After 2000 iterations these values are the same and the learning process is finished.

In Figure 7 one can see the second input signal, ε_2^p, and the coordinator parameter, γ^p. Dynamic characteristics are different. At the beginning, the difference between feedback and coordinator signals is bigger. While the coordinator increases its value, the feedback signal decreases its value.

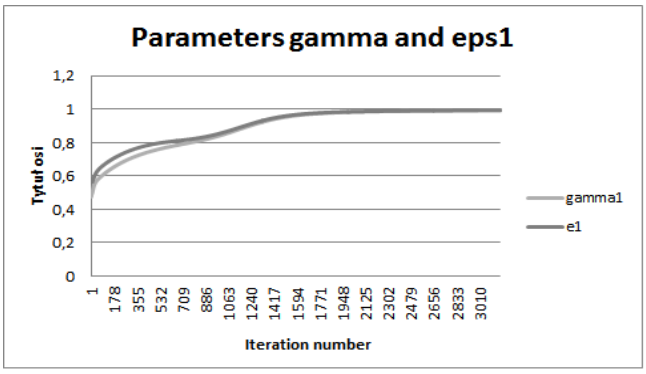

Fig. 6. Coordinator input value ε_1 and output value γ gamma- coordinator signal from coordinator to subnetworks, e1 - feedback signal from subnetwork 1 to coordinator

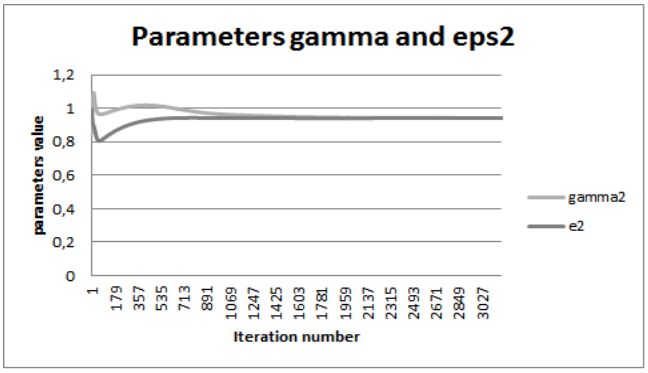

Fig. 7. Coordinator input value ε_2 and output value γ gamma- coordinator signal from coordinator to subnetworks, e2 - feedback signal from subnetwork 2 to coordinator

To summarize, one can observe that in the two-level ANN the learning processes have different dynamic characteristics. The quality of these characteristics depend not only on the input and output learning data, but also on the coordinator parameters.

Comparing the dynamic characteristics in Figures 5, 8, 9 and 10, one can notice that the characteristics are different. The main dynamic characteristics are shown in Table 1.

Table 1. The main dynamic characteristics

PARAM D	PARAM C	ITERATION	REMARK
0.04	-11.5	1600	
0.1	-4	1700	
0.2	-1.5	2100	
0.4	-0.25	1700	oscillations

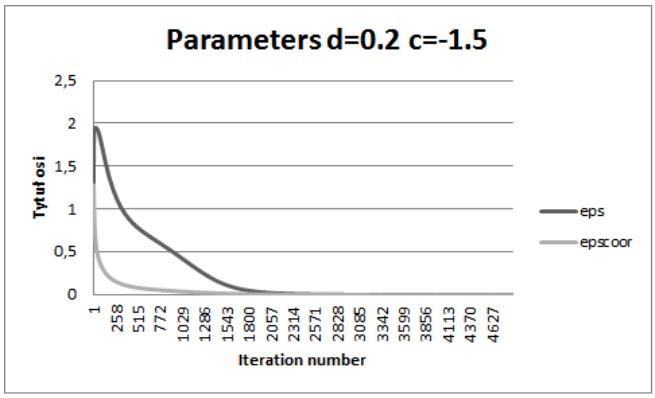

Fig. 8. Value of the global target function ϕ (eps) and coordinator target function ψ (epscoor), depending on the iteration number and parameters above

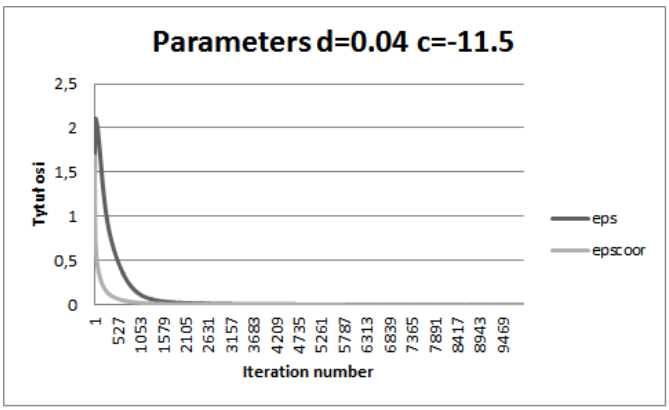

Fig. 9. Value of the global target function ϕ (eps) and coordinator target function ψ (epscoor), depending on the iteration number and parameters above

When the feedback parameter "C" decreases its value and the forward parameter "D" increases, the learning process is less stable. In Figure 10, small oscillations are seen. In the off-line learning process this kind of aspect is less dangerous than for the on-line learning process. In the on-line processes oscillations are debarred in many situations.

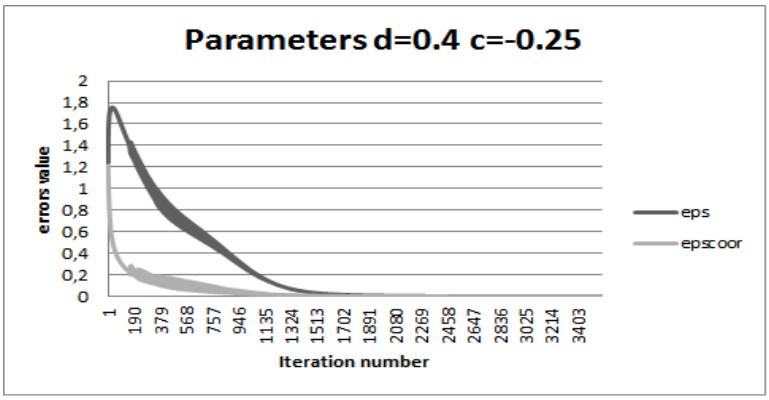

Fig. 10. Value of the global target function ϕ (eps) and coordinator target function ψ (epscoor), depending on the iteration number and parameters above

5. Conclusion

In the article the main emphasis was put on the coordination process and the coordinator structure. The coordinator has the simplest structure "P" – proportional with the internal feedback. In regard to quality and quantity, the main dynamic characteristics in the learning processes for the coordinator and all ANNs (the global target function) were shown. As it was shown, the coordinator's impact on the learning process is relatively limited. The coordinator only has to use the current information by feedback signals ε_1^p and ε_2^p. This information is insufficient to optimally coordinate achieving non-linear characteristics of the first-level sub-processes. It is obvious that for the big set of input data and the more complicated ANN structures the learning process could be more time-consuming and sophisticated. In the classification process of the non-linear separable data, a two-layer network may not be sufficient to solve this task. Further study should be conducted with respect to two directions: the implementation of the coordinator into two and more hidden layers or studying more complicated structures of the regulator, "PI" and "PID". In these regulators, additional information is used – the first difference error or the sum of errors. The quality of the coordination process should be more optimal.

References

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

[2] S. Placzek, "Decomposition and the principle of interaction prediction in hierarchical structure of learning algorithm of ANN", Poznań University of Technology, Academic Journal, *Electrical Engineering* no 84, Poznań 2015.

[3] Ch. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, Science + Business Media, LLC 2006.
[4] W. Findeisen, J. Szymanowski, A. Wierzbicki, *Teoria i metody obliczeniowe optymalizacji*. Państwowe Wydawnictwo Naukowe, Warsaw 1977.
[5] 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, No. 2, March 2007.
[6] S. Płaczek, "A two-level on-line learning algorithm of Artificial Neural Network with forward connections". Proc. Science and Information Conference 2014, London, ISBN: 978-0-9893193-1-7.
[7] Bubnicki Z., *Teoria i algorytmy sterowania*, Wydawnictwo Naukowe PWN, Warszawa 2005.

(Received: 25. 09. 2015, revised: 10. 12. 2015)