POZNAN UNIVERSITY OF TECHNOLOGY ACADEMIC JOURNALS
No 84 Electrical Engineering 2015

Stanisław PŁACZEK*

DECOMPOSITION AND THE PRINCIPLE OF INTERACTION PREDICTION IN HIERARCHICAL STRUCTURE OF LEARNING ALGORITHM OF ANN

For the most popular ANN structure with one hidden layer, decomposition is done into two sub-networks. These sub-networks form the first level of the hierarchical structure. On the second level, the coordinator is working with its own target function. In the hierarchical systems theory three coordination strategies are defined. For the ANN learning algorithm the most appropriate is the coordination by the principle of interaction prediction. Implementing an off-line algorithm in all sub-networks makes the process of weight coefficient modification more stable. In the article, the quality and quantity characteristics of a coordination algorithm and the result of the learning algorithm for all sub-networks are shown. Consequently, the primary ANN achieves the global minimum during the learning process.

KEYWORDS: Artificial Neural Network, hierarchy, decomposition, coordination, coordination principle.

1. CALCULATION COMPLEXITY

For a multi-layered ANN a lot of hidden layers and an output layer can be sectioned off. 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 hierarchical structure. Now the network consists of the two of sub-networks, for each of which a local target function is defined by $\Phi = (\Phi_1,\Phi_2)$. Similar to the ANN structure decomposition, a learning algorithm using error back-propagation can be decomposed too. 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.

There is a set of optimization tasks on the first level in a learning algorithm constructed this way. 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, there are standard procedures to solve these problems.

* Akademia Finansów i Biznesu Vistula.

But in a two-level learning algorithm structure, a coordinator is not responsible for solving the global task Φ. A coordinator is obliged to calculate the value of coordination parameters $\gamma = (\gamma_1,\gamma_2)$ for every task on the first level. The first level, searching for the solution of all tasks, has to use the coordination parameters value. It is an iterative process. A coordinator in every iteration cycle receives new values of feedback parameters $\varepsilon = (\varepsilon_1,\varepsilon_2)$ from the first-level tasks. Using this information, coordinator has to make new decisions – calculate a new coordination parameters value (Fig. 1).

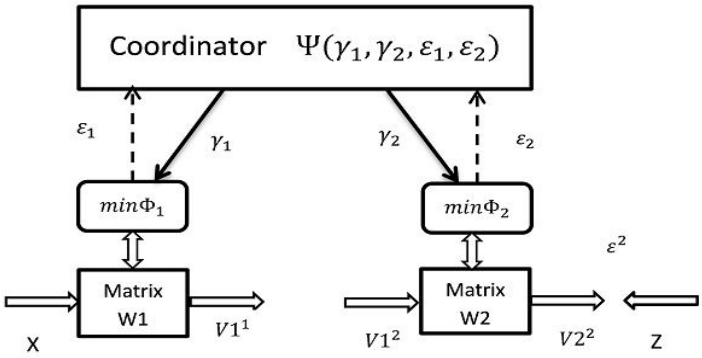

Fig. 1. Decomposition and coordination scheme

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

- the global target function Φ,
- two local target functions Φ_1, Φ_2,
- the coordinator target function ψ.

According to [1], the solution of the first-level tasks should be built in such a way that when all the first-level tasks are solved, the final solution must be achieved – the minimum of the global target function. This kind of stratified structure is known as the level hierarchy [1, 6].

2. COORDINATION BY INTERACTION PREDICTION

In a two-level learning algorithm, the coordinator plays the main role. It is now time to decide what kind of coordination principle will be chosen. This principle specifies various strategies for the coordinator and determines the structure of the coordinator. In [1], three ways were introduced in which the interaction could be performed.

- Interaction Prediction. The coordination input may involve a prediction of the interface between the first and the second sub-network (Fig. 1). For the first

sub-network, the coordinator signal $\gamma = \gamma_1$ is treated as an expected output vector given by the teacher. It is being sent into the target function Φ_1 as a parameter. For the second sub-network, the coordinator signal $\gamma = \gamma_2$ is treated as an input vector of the second sub-network. This vector in the target function Φ_2 is treated as an input variable. So, in the first level of sub-networks coordination interface plays a different role. For the first sub-network it is the teacher's parameter, but for the second sub-network – the input variable. Both tasks are fully specified and algorithms have the ability to find the minimum value of their target functions Φ_1, Φ_2. The other two coordination principles are defined as:

- Interaction Decoupling. Each first-level sub-system is introduced into the solution of its own task and can treat the interface input as an additional decision variable to be free. This means that sub-systems are completely decoupled.
- Interaction Estimation. The coordinator specifies the ranges of interface inputs over which they may vary.

As stated above, the coordinator needs the feedback information from the first-level sub-networks, checking if the predicted signals $\gamma = (\gamma_1,\gamma_2)$ were true. If not, the coordinator – using its own target function – should find a new value of the coordination signal. The first sub-network calculates a new value of its output signal which, at the same time, is the feedback signal ε_1 to the coordinator. The second sub-network is trying to minimize the local target function Φ_2 and calculate a new optimal value of input signal ε_2 which is being sent to the coordinator. Therefore, the coordinator has full information and is ready to calculate and predict the new value of the coordination input signal γ_1, γ_2. Taking into account that

$$\gamma = \gamma_1 = \gamma_2 \tag{1}$$

the coordinator target function is defined as:

$$\psi = \frac{1}{2}\sum_{p=1}^{N_p}\sum_{i=1}^{N_1}(\varepsilon 1_i^p - \gamma_i^p)^2 + \frac{1}{2}\sum_{p=1}^{N_p}\sum_{i=1}^{N_1}(\gamma_i^p - \varepsilon 2_i^p)^2 \tag{2}$$

Using gradient algorithm one can calculate:

$$\frac{\partial\psi}{\partial\gamma_i^p} = 2 * \gamma_i^p - (\varepsilon 1_i^p - \varepsilon 2_i^p) \tag{3}$$

The new value of the coordinator signals $\gamma = \gamma_1 = \gamma_2$

$$\gamma_i^p(n+1) = \gamma_i^p(n) - \lambda_1 * \frac{\partial\psi}{\partial\gamma_i^p} \tag{4}$$

where: λ_1 – a learning coefficient for the coordinator iteration process.

3. TWO-LEVEL LEARNING ALGORITHM STRUCTURE

In the decomposed ANN structure one can define the subsequent target functions:

- The global target function. For all epoch:

$$\psi(W1, W2, X, Y) = \frac{1}{2}\sum_{k=1}^{N_2}\sum_{p=1}^{N_p}(y_k^o - z_k^p)^2 \tag{5}$$

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:

$$\phi_1(W1, X, \gamma) = \frac{1}{2}\sum_{p=1}^{N_p}\sum_{i=1}^{N_1}(\nu 1_i^p - \gamma_i^p)^2 = \frac{1}{2}\sum_{p=1}^{N_p}\sum_{i=1}^{N_1}(f(\sum_{j=0}^{N_0} W1_{ij} * x_j) - \gamma_i^p)^2 \tag{6}$$

where: γ[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:

$$\phi_2(W2, Z, \gamma) = \frac{1}{2}\sum_{p=1}^{N_p}\sum_{k=1}^{N_2}(\nu 2_k^p - z_k^p)^2 = \frac{1}{2}\sum_{p=1}^{N_p}\sum_{k=1}^{N_2}(f(\sum_{i=0}^{N_1} W2_{ki} * \gamma_k^p) - z_k^p)^2 \tag{7}$$

where: γ[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 (6), one can calculate the feedback signal $\varepsilon 1_i^p$ and the new value of matrix W1.

$$\varepsilon 1_i^p = f(\sum_{j=0}^{N_0} W1_{ij} * x_j) \tag{8}$$

$$\frac{\partial \phi_1}{\partial W1_{ij}} = \sum_{p=1}^{N_p}(\nu 1_i^p - \gamma_i^p) * f' * x_j^p \tag{9}$$

$$W1_{ij}(n+1) = W1_{ij}(n) - \alpha_1 * \frac{\partial \phi_1}{\partial W1_{ij}} \tag{10}$$

For the second sub-network using (7), one can calculate the new value of $\varepsilon 2_i^p$ and the new value of matrix $W2_{ki}$.

$$\frac{\partial \phi_2}{\partial W2_{ki}} = \sum_{p=1}^{N_p}(\nu 2_k^p - z_k^p) * f' * \gamma_k^p \tag{11}$$

$$W2_{ki}(n+1) = W2_{ki}(n) - \alpha_2 * \frac{\partial \phi_2}{\partial W2_{ki}} \tag{12}$$

$$\frac{\partial \phi_2}{\partial \gamma_i^p} = \sum_{k=1}^{N_1} (\nu 2_k^p - z_k^p) * f_2^{'} * W2_{ki} \tag{13}$$

$$\varepsilon 2_i^p(n+1) = \gamma 2_i^p(n) - \alpha_3 * \frac{\partial \phi_2}{\partial \gamma 2_i^p} \tag{14}$$

4. EXAMPLE

In the example the main dynamic characteristics of the learning process are shown. The emphasis is placed on the characteristics of the first-level local target functions Φ_1, Φ_2 and when it comes to the second-level, the coordinator target function ψ. The structure of ANN is simple and can be described as ANN(4-5-1). This means that the ANN includes: 4 input neurons, 5 neurons in a hidden layer and 1 output neuron. Sigmoid activation functions are implemented in both the hidden and the output layers. Four arguments of XOR4 function are fed as input data. Consequently, every epoch includes 16 vectors. By changing different learning parameters as, for example, α_1, α_2, α_3, λ_1, dynamic characteristics have been studied. Fig. 2 and 3 show how the two target functions Φ_1, Φ_2 changed their values during the learning process (iterations' number).

The quality of the dynamic processes is different. The function Φ_2 represents the second local target function (output one). This process is smooth which means that during the learning process the value of Φ_2 decreases its value monotonically.

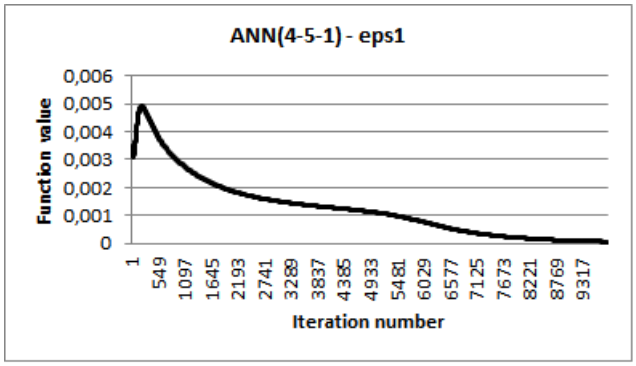

Fig. 2. Value of the local target function Φ_1 depending on iteration number

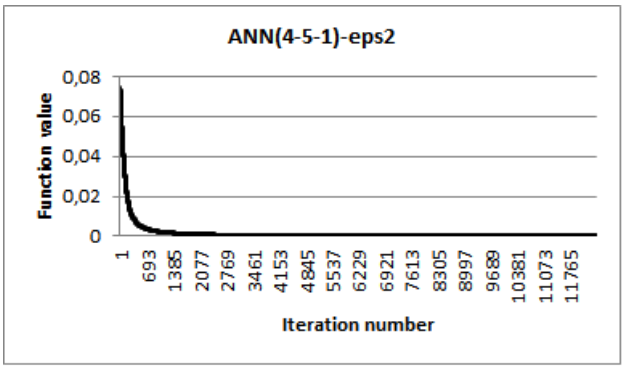

Fig. 3. Value of the local target function Φ_2 depending on iteration number

The quality of dynamic processes Φ_1 is different. At the beginning, Φ_1 increases its value and after 300 iterations decreases its value in a monotonic way. As one can observe it in the second sub-network, this process does not decrease its value. Between 2000 and 5000 iterations, the learning process decreases its value very slowly.

After that, the learning process monotonically achieves its minimum value and the whole learning process is finished. This can be explained by comparing the dimensionality of two matrices W2 and W1. Matrix W1 (5x5) contains 25 neurons while matrix W2 (6x1) contains only 6. Matrix W1 needs more time (more iteration) to change all the neurons' value using learning data encoding by vector γ. This value is calculated by the coordinator, which has to use the ε_2 feedback value from the second sub-network. The learning process is very dynamic and the information between the coordinator and the two sub-networks is changed. At the end of the learning process, the differences between the coordinator value γ and the feedback signals are very small and the learning process achieves the final target values Φ_1, Φ_2 very slowly.

To coordinate the sub-networks, using a simple interaction prediction algorithm is not optimal, especially in the final part of the learning process. Probably the PID regulator as the coordination algorithm will be more effective. This study should be continued.

In Fig. 4, the learning characteristic of the global target function is shown. Using a simple interpretation one may say that it is a sum of the two local target functions Φ_1, Φ_2. However, this is not true because both target functions are non-linear from the weight coefficient point of view. Thus, a simple additive principle could not be applicable. Finally, the characteristic of the coordinator target function is shown in Fig. 5. It is smooth and achieves a small value after 2000 iterations.

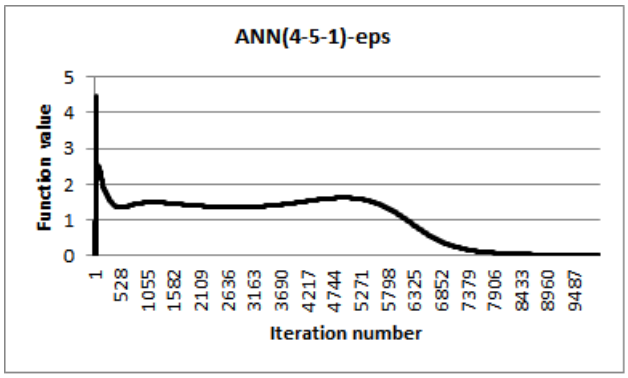

Fig. 4. Value of the global target function Φ depending on iteration number

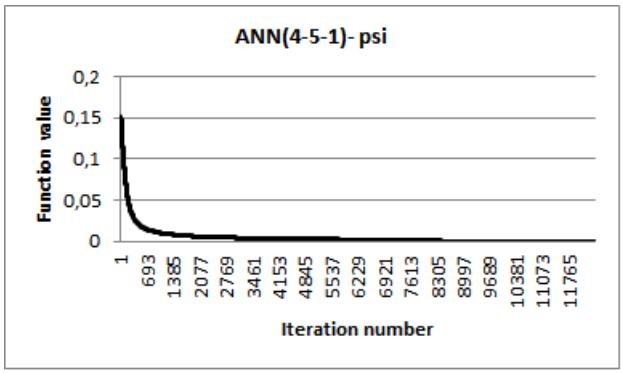

Fig. 5. Value of the coordinator target function ψ depending on iteration number

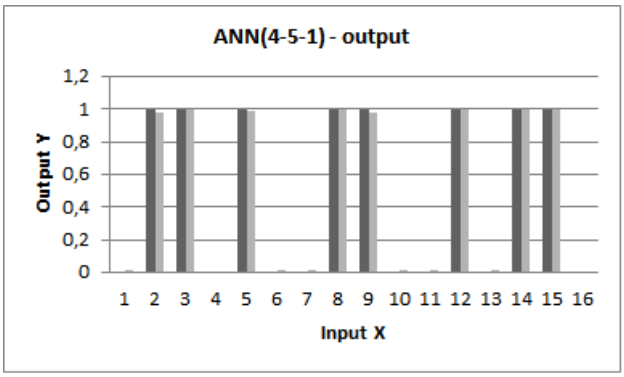

Fig. 6. Result of the ANN's learning

The learning result is shown in Fig. 6. The quality is very good and the hierarchical learning algorithm works perfectly.

CONCLUSION

The coordination algorithm is not optimal from the iteration point of view and needs a lot of iterations to achieve the final result, namely finishing the learning process. In future work the second coordination principle should be tested - Interaction Decoupling.

REFERENCES

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

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

[3] W. Findeisen, J. Szymanowski, A. Wierzbicki, *Teoria i metody obliczeniowe optymalizacji,* Państwowe Wydawnictwo Naukowe, Warsaw 1977.

[4] 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.

[5] S. Placzek, "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.

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