ELEMENTS OF AN ALGORITHM FOR OPTIMIZING A PARAMETER-STRUCTURAL NEURAL NETWORK

Maria Mrówczyńska

Department of Land and Environment Engineering, Institute of Building Engineering
University of Zielona Góra, Poland

Abstract

The field of processing information provided by measurement results is one of the most important components of geodetic technologies. The dynamic development of this field improves classic algorithms for numerical calculations in the aspect of analytical solutions that are difficult to achieve. Algorithms based on artificial intelligence in the form of artificial neural networks, including the topology of connections between neurons have become an important instrument connected to the problem of processing and modelling processes. This concept results from the integration of neural networks and parameter optimization methods and makes it possible to avoid the necessity to arbitrarily define the structure of a network. This kind of extension of the training process is exemplified by the algorithm called the Group Method of Data Handling (GMDH), which belongs to the class of evolutionary algorithms. The article presents a GMDH type network, used for modelling deformations of the geometrical axis of a steel chimney during its operation.

Keywords: *neural networks, group data handling, deviations from the vertical.*

1. Introduction

Artificial neural networks are used for solving a number of problems in different fields of science and technology because of their easy implementation and ability to approximate functions without the necessity to describe the relationship between input and output data. However, neural networks are liable to a method error, which is difficult to define and eliminate, resulting from the fact that the structure of the network is adopted arbitrarily. When neural networks are used for modelling complex problems this error has a significant influence on the evaluation of the final training result. Therefore, it is suggested that the teaching process should also be extended onto the connections between neurons. An example of this kind of solution can be the Group Method of Data Handling – GMDH, the concept of which was prepared by

Iwachnienko (Iwachnienko, 1982; Duch et al. 2000). The GMDH network belongs to the group of self-organizing networks. It is hierarchical in structure created automatically on the basis of teaching (training) and testing data sets, is a, which consists of polynomial partial models (Korbicz, 2009). The result of the operation of the GMDH procedure is the Iwachnienko polynomial, which ensures the high accuracy of results obtained during the operation of the network as well as the practical usefulness of this solution (Duch et al. 2000).

   In the article the Group Method of Data Handling is used for modelling deformations of the geometrical axis of a steel chimney for venting subterranean gas tanks on the basis of cyclic measurements started in 2007.

2. GMDH Algorithm

Errors resulting from the arbitrary adoption of network architecture at the design stage can be minimized by merging the teaching process with the determination of an optimum structure of the neural network. The Group Method of Data Handling consists of replacing the whole model of the neural network with a hierarchical structure, which is built of polynomial partial models. The network itself is constructed by merging a specific number of single neurons n (Fig. 1) processing the input signal **x** into the output signal y according to the dependence described by the transfer function f:

$$y = f(\mathbf{x}) = f(x_1, x_2, \ldots, x_m),$$
(1)

with the assumption that at least two input signals from among all possible ones $x_1, x_2, \ldots x_m$ are stimulations.
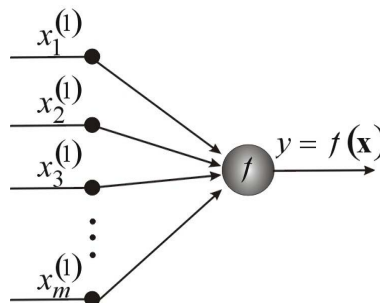
Fig. 1. The structure of a GMDH type neuron (prepared by the author)

   The algorithm permits different variants of the transfer function f, which is non-linear in most cases, and its precise definition is of little significance. It is most important that the transfer function should meet the conditions determined by the dependence:

$$y = f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2),$$
(2)

which is the Nth-rate approximation of the Kołmogorow-Gabor polynomial defined as (Iwachnienko, 1971):

$$y = a_0 + \sum_{i=1}^{N} a_i x_i + \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij} x_i x_j + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} a_{ijk} x_i x_j x_k + \ldots,$$
(3)

where $a_0, a_i, a_{ij}, a_{ijk}$ are parameters of the polynomial.

The transfer function (1) should not be too expanded because it would make the teaching time longer, complicate the teaching process itself and make it impossible to accurately assess the teaching error. In order to meet the requirements that the transfer function should have a simple form and the approximation should agree with the form described by formula (2), it is assumed that $N=2$. Then, the polynomial (3) assumes the form:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_{11} x_1^2 + a_{22} x_2^2 + a_{12} x_1 x_2.$$

(4)

The synthesis of a GMDH network consists in estimating parameters of particular partial models iteratively and merging them by means of adequately chosen selection models (Fig. 2). This course of action leads to the final resultant structure of the network that makes it possible to obtain the best output signal according to the adopted identification criterion.
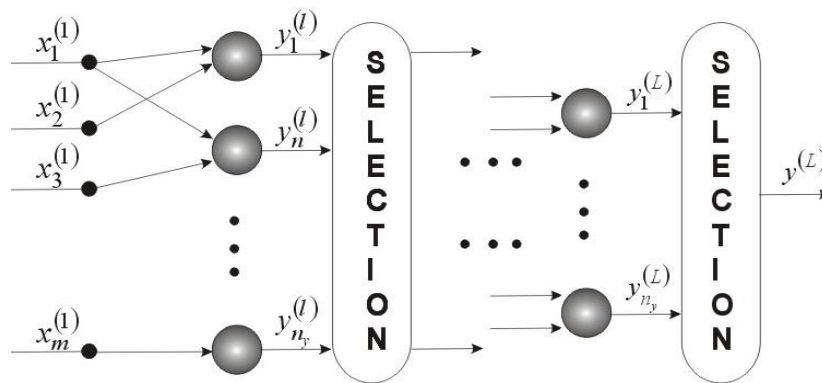
Fig. 2. The concept for the synthesis of a GMDH type network
(prepared by the author, based on Korbicz, 2009)

As has been mentioned earlier, the teaching process is iterative and leads to the evolution of the resultant structure of the network. During the first iteration an input layer of neurons described by the activation function (1) is built, including all combinations of the input signals x_i. Thus, it is possible to create k new elements whose number is determined by the formula for the number of $k\,z\,m$ combinations of elements after n. It is assumed that $n<m$

$$k = \frac{m!}{n!(m-n)!}$$

(5)

where: m – the number of input signals, n – the number of input neurons. If the activation function is defined in the form of a 2nd degree polynomial (4), then the number of neurons in the output layer is expressed by formula:

$$s = \frac{m(m-1)}{2}.$$

(6)

The neurons in this layer (Fig. 3) are described by the activation function in the form:

$$y_1^{(1)} = f\left(x_1, x_2, \mathbf{a}_{12}^{(1)}\right)$$

$$\ldots$$

$$y_{m-1}^{(1)} = f\left(x_1, x_m, \mathbf{a}_{1m}^{(1)}\right)$$

$$y_m^{(1)} = f\left(x_2, x_3, \mathbf{a}_{13}^{(1)}\right)$$

$$\ldots$$

$$y_{2m-3}^{(1)} = f\left(x_2, x_m, \mathbf{a}_{2m}^{(1)}\right) \tag{7}$$

$$\ldots$$

$$y_k^{(1)} = f\left(x_{m-1}, x_m, \mathbf{a}_{m-1,m}^{(1)}\right)$$

where the unknown parameters $\mathbf{a}_{12}^{(1)}, \ldots, \mathbf{a}_{m-1,m}^{(1)}$ are optimized for each neuron by means of the least square method or a different teaching method e.g. the Widrow-Hopf delta rule (Duch et al., 2000).
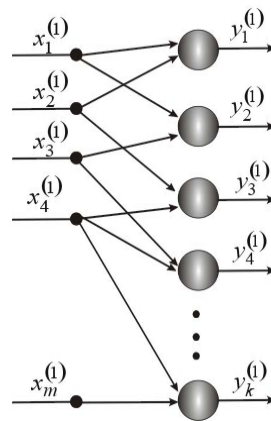

Fig. 3. The output layer of a GMDH network
(prepared by the author, based on Duch et al., 2000)

Another step consists in attaching the newly created layer to the network. This action is preceded by the selection of the neurons that process input signals most accurately. The selection process is intended to eliminate the elements whose processing error $E(y)$ is too big according to the adopted criterion. It is possible to distinguish the following selection methods (Duch et al,. 2000):

– the method of permanent populations,
– the method of optimum populations,
– the method of decreasing populations.

All the above mentioned selection procedures have been used in this article, and the best results were obtained with the use of the method of optimum populations, which consists in rejecting the neurons whose processing error $E(y)$ reaches values higher than the arbitrarily determined threshold ε. Figure 4 presents the neuron selection process for the method of optimum populations.
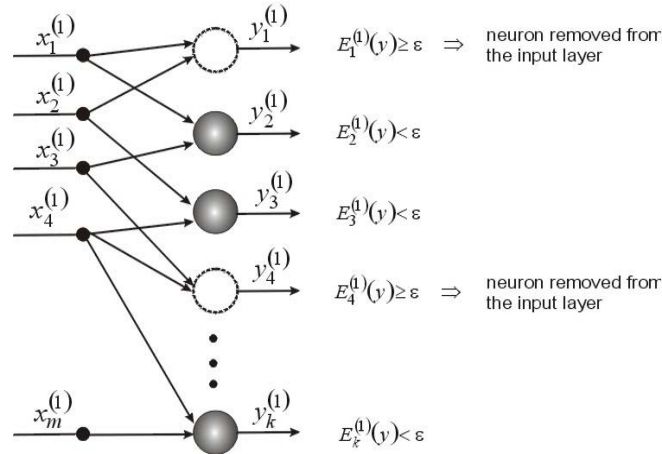
Fig. 4. The selection of neurons in the input layer
(prepared by the author, based on Duch et al., 2000)

During the second iteration, the output signals from the previous layers are used as input data. In the simplest case of two input signals for each neuron, the output signal of the neurons in the layer l can be written as

$$y_1^{(l)} = f\left(y_1^{(l-1)}, y_2^{(l-1)}, \mathbf{a}_{12}^{(l-1)}\right),$$

$$\ldots$$

$$y_{k-1}^{(l)} = f\left(y_1^{(l-1)}, y_k^{(l-1)}, \mathbf{a}_{1k}^{(l-1)}\right),$$

$$y_k^{(l)} = f\left(y_2^{(l-1)}, y_3^{(l-1)}, \mathbf{a}_{23}^{(l-1)}\right),$$

$$\ldots$$

$$y_{2k-3}^{(l)} = f\left(y_2^{(l-1)}, y_k^{(l-1)}, \mathbf{a}_{2m}^{(l-1)}\right),$$

$$\ldots$$

$$y_t^{(l)} = f\left(y_{k-1}^{(l-1)}, y_k^{(l-1)}, \mathbf{a}_{k-1,k}^{(l-1)}\right) \tag{8}$$

The network architecture in further layers is created in the same way until the optimality criterion Q_{opt} is met (Luzar, 2010). From among a number of definitions of the function of the optimality criterion, the one used in this article is the convergence criterion defined as (Duch et al., 2000):

$$Q_{opt} = \frac{v}{m_v} \tag{9}$$

where: v – the correction, m_v – the error of correction. The optimality criterion Q_{opt} makes it possible to determine the processing error of a single neuron, which is the basis for making the decision whether it should be included in the following layer or rejected. The whole synthesis process of the GMDH network is controlled by reference to external data, which do not participate in the teaching process i.e. to the data included in the test set. If the number of data is sufficiently great, it is also possible to separate a control set whose task is to obtain a better final solution. According to Gödel's theorem (Duch et al., 2000), the control data are the basis for determining the criterion for the operation of the algorithm, independently of the criteria used at the network synthesis stage.

3. Numerical example

The dimensions of tall tower structures are disproportionate, their height is much greater than the dimensions of their cross sections. Industrial steel chimneys are light, inexpensive and easy to build but because of their slenderness, they have to be secured by anchoring cables or lattice structures (Gocał, 2010). The location of the geometrical axis of a steel chimney with a height of 80m was determined with the method of surrounding tangents. The directions tangent to the 7 determined cross sections of the chimney were observed from the point of the measurement control network with the use of a total station with the nominal accuracy of direction measurement $m_k = 20^{cc}$. The measurement was carried out for two positions of the telescope in relation to the sides of the measurement control network.

The numerical realization of the task consisted in determining the values of the unknowns:

- the coordinates of the centres of the sections (x_s, y_s),
- the length of the radius r in the observed cross sections,

taking into considerations the accuracy characteristics of these parameters. The values of the unknowns were determined with the least square method on the basis of a system of approximation equations for each observed cross section. For each tangent direction, the approximation equation has the form (Czaja, 1983):

$$v_r = (\sin A)\mathrm{d}x_s - (\cos A)\mathrm{d}y_s \pm \mathrm{d}r + l \tag{10}$$

and the absolute term calculated on the basis of the approximated values of the parameters:

$$l = \left(x_s^0 - X\right)\sin A - \left(y_s^0 - Y\right)\cos A \pm r_0 \tag{11}$$

The symbols in formula (10) and formula (11) denote:

- v_r – the correction to the observation,
- A – the bearing of the tangent direction,
- x_s^0, y_s^0 – the approximate coordinates of the geometrical centre of the cross section of the observation,
- r_0 – the approximate value of the radius length,
- X, Y – the coordinates of the point of the measurement control network from which the tangent direction was observed.

The analysis comprised the results of the measurements of deviations from the vertical of the geometrical axis of the chimney carried out in the years 2009-2011. Optimum network architecture was built with the use of the GMDH algorithm, which made it possible to obtain results in the form of deviations from the vertical of the geometrical axis of the chimney in relation to the increments in the height of the observation levels. As has been mentioned before, the construction process of the network structure is carried out until the processing error $E(y)$ begins to increase (Fig. 5), then we assume that the network architecture is optimum. Figure 6 presents the results of the operation of the Group Model of Data Handling algorithm with the use of the method of optimum populations as the selection method, which has the

smallest processing error $E(y)$=6.7 mm as well as the other methods: the method of permanent populations, (the processing error is $E(y)$=10.6 mm) and the method of decreasing populations (the processing error is $E(y)$=11.4 mm).
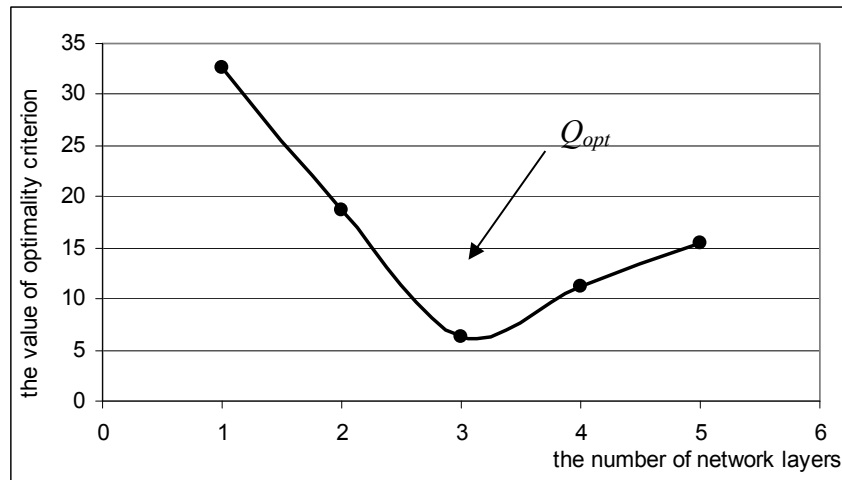
Fig. 5. The deviations obtained from the GMDH algorithm (the method of optimum populations, prepared by the author)
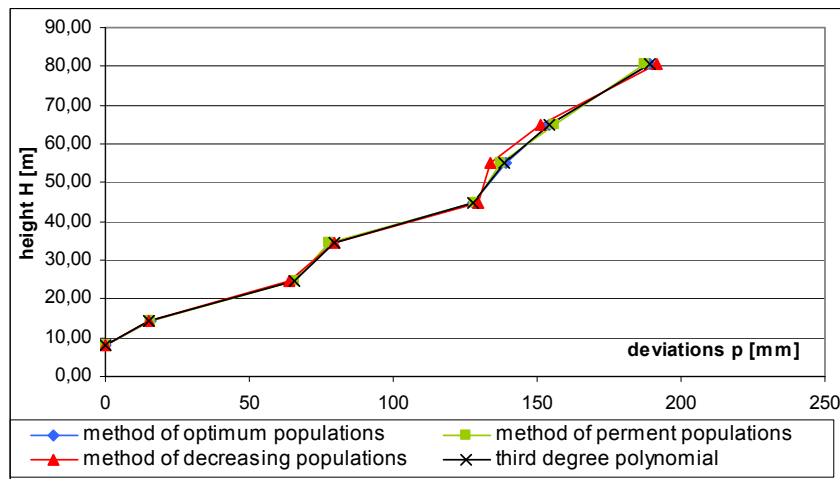
Fig. 6. The results of the operation of the algorithm with the use of particular selection methods (prepared by the author)

Tab. 1. The results of the approximation with the use of the GMDH (prepared by the author)

The height of the observation level H [m]	Deviations from plumb p [mm]			
	the method of optimum populations	the method of permanent populations	the method of decreasing populations	the third degree polynomial
8,25	0,0	0,0	0,0	0,0
14,36	15,5	15,4	15,0	15,3
24,49	65,3	65,5	64,0	65,4
34,63	79,2	77,6	79,5	79,2
44,78	127,7	128,4	129,4	127,8
54,93	138,9	137,3	133,6	138,8
65,08	154,2	156,1	151,2	154,2
80,72	189,0	187,6	191,3	189,0

The results of the approximation with the use of the Group Method of Data Handling are presented in detail in Table 1, and they are compared with the results obtained by approximation with a third degree polynomial. The results obtained with the use of the method of optimum populations as the selection method (Fig. 6) are closest to the results obtained with the use of a third degree polynomial. Figure 7 presents the differences between the results obtained with the use of classic calculation methods and the results obtained with the use of the GMDH algorithm for the selection methods discussed.
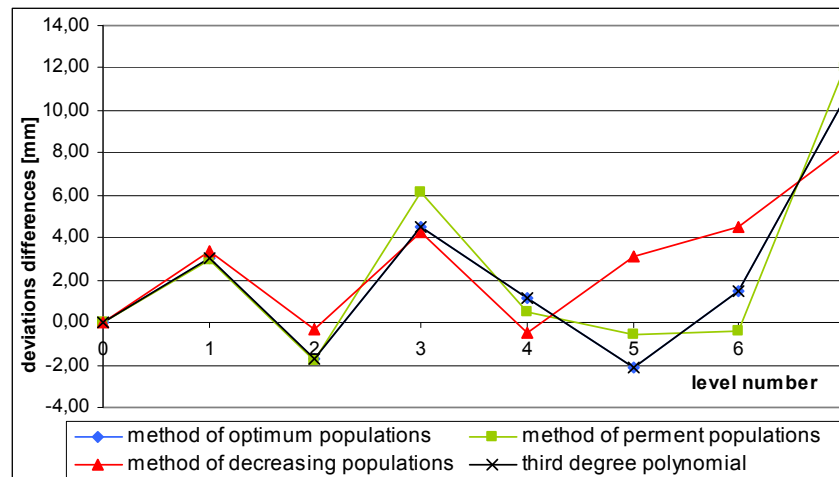

Fig. 7. The differences between the values of deviation for the adopted selection methods (prepared by the author)

4. Conclusions

The group data handling algorithm presented in the article makes it possible to solve problems that restrict classic neural networks. The presented course of action does not include an arbitrarily defined network structure but it solves the problem of determining its compact topology as an additional teaching result. An optimum structure (for a particular problem) is built during the teaching process, which makes it possible to increase its effectiveness. The structure construction procedure is continued as long as its development leads to increasing the effectiveness of processing experimental data.

The algorithm discussed in this article is one of the intelligent calculation systems because of its ability to teach neural networks and to optimize the parameters of constant transfer functions. The user may choose the form of the function, the criterion functions and the methods for selecting neurons. Bearing in mind the abovementioned characteristics of the algorithm, the presented solution represents the structural-parameter optimization of a neural network.

Apart from the abovementioned advantages of the GMDH algorithm, it is worth mentioning its characteristics such as elements of the theory of genetic algorithms widely used for solving optimization problems. Formula (2) says that the dependences $y = f_1(x_1)$ and $y = f_2(x_2)$ are the general notations of the mutation function, and the dependence $y = f_{12}(x_1, x_2)$ is the general notation of the crossing function. Therefore, the presence of elements of the theory of genetic algorithms in the GMDH algorithm is important from the point of view of the procedure of the structural-parameter network optimization.

Due to these characteristics the presented solution is universal and can be used in a number of fields when there are small and limited data sets.

References

Czaja J. (1983). *Geodezja inżynieryjno – przemysłowa*. Skrypt Uczelniany nr 893 Akademii Górniczo – Hutniczej w Krakowie. Kraków.

Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R. (2000). *Sieci neuronowe*. Biocybernetyka i inżynieria biomedyczna, tom 6. Akademicka Oficyna Wydawnicza EXIT. Warszawa.

Gocał J. (2010). *Geodezja inżynieryjno – przemysłowa*, tom III. Wydawnictwo Akademii Górniczo – Hutniczej. Kraków.

Iwachnienko A.G., (1971). Polynominal theory of complex systems. *IEEE Trans. Systems, Man and Cybernetics*, Vol.SMC-1, No.4.

Iwachnienko A.G., (1982). *Induktiwnyj metod samoorganizacji modelej słożonych system.* Naukowa Dumka. Kijów.

Korbicz J. (2009). Sztuczne sieci neuronowe i ich zastosowanie w elektrotechnice i energetyce. *Przegląd Elektroniczny* R. 85, Nr 9. Warszawa.

Luzar M. (2010). Przybornik modelowania neuronowego GMDH. *Conference Archives PTETiS* Vol. 28.

Authors:

Maria Mrówczyńska [1], m.mrowczynska@ib.uz.zgora.pl
[1] Department of Land and Environment Engineering,
Institute of Building Engineering,
University of Zielona Góra,
ul. Prof. Z. Szafrana 1, 65-516 Zielona Góra, Poland