# Comparing NN and GMDH methods for prediction of socio-economic processes

Bulgakova Oleksandra, Samoilenko Oleksandr

*International Research and Training Center of Information Technologies and Systems of NAS and MES of Ukraine, Prospekt Akademika Glushkova 40, Kyiv, 03680,Ukraine*

sashabulgakova@list.ru, astrid@irtc.org.ua

**Abstract.** *There are considered two different methods for prediction of socio-economic processes: the combinatorial GMDH algorithm and an artificial neural network. These methods are analyzed in the task of modeling of the Ukrainian gross domestic product (GDP) as dependent from input arguments (investments). An analysis and comparison of these methods showed interesting results that gives pre-conditions to use capabilities of neural networks jointly with the GMDH algorithms.*

## Keywords

Artificial neural networks, Group Method of Data Handling (GMDH), linear neural networks, method of least squares.

## 1 Introduction

Scientific and technical progress deeply influences on development of economy. The efficiency and quality of the made decisions which are adopted by the leaders of different levels considerably depends on the timely providing of process management necessary reliable information, which characterizes processes and phenomena, what be going with the object of management. In general administrative cycle prediction is only one of components but the value of this component can not be underestimated.

The theory of artificial neural networks develops quickly to be support for the most optimistic projects. The today's explosion of interest brought over thousands of researchers to the neural networks especially for prediction.

## 2 The problem description

It is necessary to build a model to predict dependence of GDP on the values of the chosen arguments. These arguments are as follows: 1) investments volume in the fixed assets in actual prices; 2) real increase (decline) of investments in the fixed assets; 3) degree of capital assets wear; 4) pure increase of direct external investments; 5) a percentage of foreign investments in the total volume of investments; 6) total volume of investments; 7) ratio of investments volume to the cost of capital assets; 8) ratio of volumes of investments in the fixed assets to GDP; 9) ratio of net investments increase to GDP; 10) ratio of total investments volume to GDP. Two methods are used: neural networks and combinatorial GMDH algorithm to solve this task. Efficiency of these methods are compared.

## Prediction using neural networks

A program for prediction by neural networks is developed in the of MATLAB environment using the most simple linear neural network and back propagation algorithm, which determines two streams in a network.

A linear model is a network without intermediate layers which in the output layer contains only linear elements (with the linear activating function). Weights correspond to the matrix elements and thresholds to the displacement vector components. When working a network actually multiplies the input vector by the weighting matrix and then the displacement vector is added to the result.

A linear model is set by the equation:

$$Y = XW + B$$

where $W$ is a matrix of network weights and $B$ is a vector of displacement.

In the MATLAB environment there is a planning function of the new linear layer. This function of the input and output vectors determines weight of linear neural network by the least-squares method.

Back propagation algorithm determines two streams in a network: forward stream is from the input layer to the output and backward reversely. It is necessary to correct weighing coefficients in the process of learning reflected in the Fig. 1.

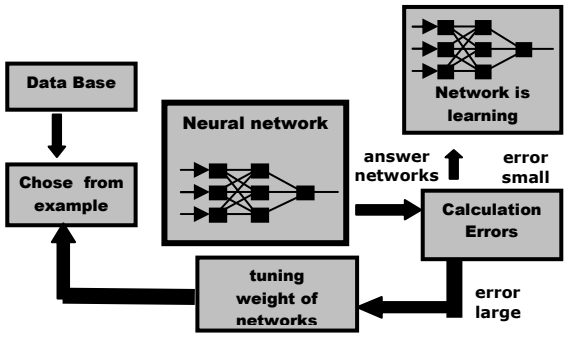


**Fig. 1.** Learning process

Using the software it is possible to build the model both with one argument (Fig. 2) and with all ones from the given sample (Fig. 4). Fig. 3 shows the GDP dependence of one argument and one can see the very large prediction error. It means that in general the prediction based on this argument will only multiply the error. This way we can define arguments negatively influencing the efficiency of models and to exclude them from the sample, so one can enhance the prediction accuracy for the general relationship (Fig. 5).

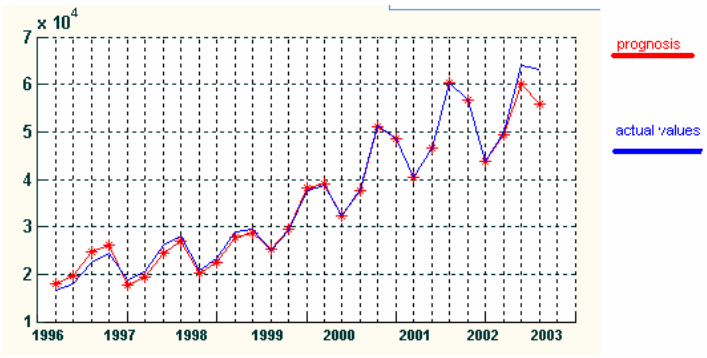


**Fig. 2.** Prediction of GDP as dependent from the investments volume in the fixed assets in actual prices

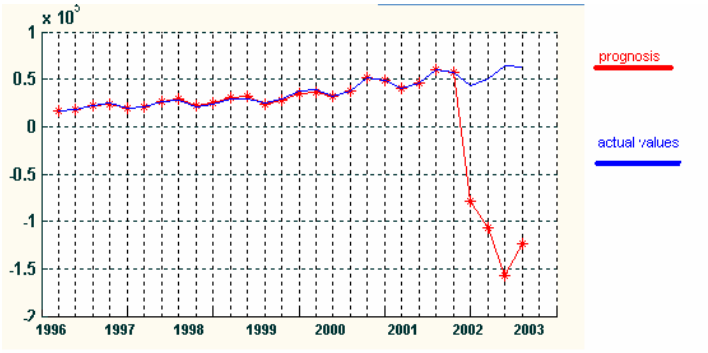


**Fig. 3.** GDP prediction as dependent from the investments volumes in the fixed assets.

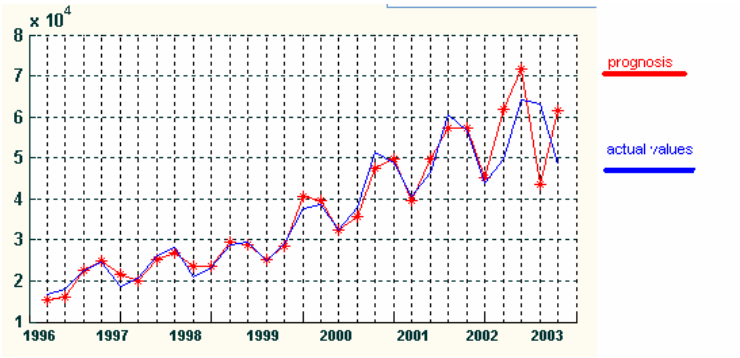


**Fig. 4.** GDP prediction from all arguments of a sample.

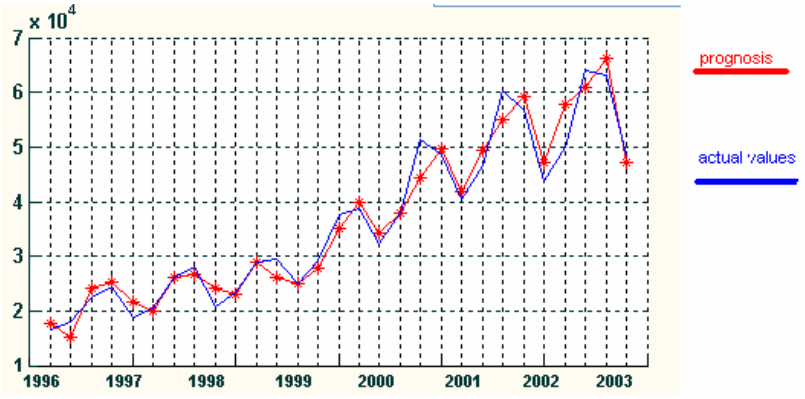


**Fig. 5.** GDP prediction based on all arguments of a sample but the argument of Fig. 2.

## Prediction based on the combinatorial GMDH algorithm

It is necessary to establish a dependence of the form $y = X_s\theta_s$, where $\theta = [\theta_1\theta_2...\theta_s]^T$ is the vector of unknown parameters being estimated. This vector determines a model of the sample.

The model quality is determined as the value of the regularity criterion *AR(s)* based on a division of the sample $X$ into 2 subsamples $X_A$ and $X_B$. We estimate arguments of the model on the training subsample $X_A$ and calculate the error on the testing subsample $X_B$:

$$AR(s) = \| y_{Bs} - X_{Bs}\theta_{As} \|^2 , \qquad (3)$$

where $\theta_{As}$ is the vector of parameters determined on the subsample $X_A$.

Two experiments are carried out by the developed software. These experiments are as follows: 1) building and analyzing the model containing all arguments of the sample; 2) building and analyzing the model based on optimal arguments selected by GMDH (Fig. 6).

As one may see from the experiments, the model of optimal complexity performs definitely better.

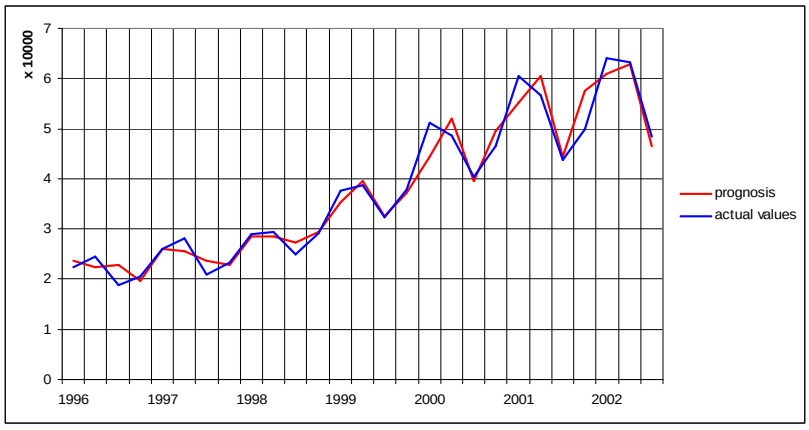


**Fig. 6.** GDP prediction by a model consisting four arguments of a sample

The results of the experiments are as follows:

For GDP prediction by NN using all arguments, the accuracy of the worst prediction for the whole period is near 90% and the best one is almost 94%.

For the case of the prediction using GMDH, the respective figures are 80% (worst) 90% (best).

# 3 Conclusion

The conducted researches displayed the following:

1) models in both methods perform better with the use of more informative arguments;
2) neural networks show slightly better results as comparing to the combinatorial GMDH algorithm.

These results give good pre-conditions to using capabilities of neural networks jointly with the GMDH algorithms.

# 4 References


[1] Dyakonov V.P., Kruglov V.V. – MATLAB 6.5 SP1 7/7 SP1/7 SP1 + Simulink 5/6. The instruments of artificial intelligence and bioinformatics. – Moskow.: SOLON - PRESS, 2006. – 456 p. (In Russian)

[2] Korotky S. Neural networks: fundamental regulations.

[3] Zaychenko Y.P. Foundations of the design of intellectual systems. – Kyiv, 2003. (In Ukrainian)