Model-Free Analogues As Active Neurons For Neural Network Self-Organization

Ivakhnenko G.A.

Survey

1. Analogues Complexing GMDH algorithm

Analogues Complexing (AC) algorithm is one of the spectrum of Group Method of Data Handling algorithms. It is used for forecasting, extrapolation and pattern recognition of ill-defined objects. In the case of insufficient a priori information, not very accurate measurements, noisy and short data sample, better results are reached by search of analogues in prehistory and by construction of non-physical models using this analogues. Such models can be considered as discrete form of physical model of ill-defined objects.

Almost all objects for recognition and control in economics, ecology, biology and medicine are undeterministic or fuzzy. Deterministic (robust) part and additional black boxes acting on each output of object can represent them. The only information about these boxes is that they have limited values of output variables, which are similar to the corresponding states of object. According to Ashby [2] diversity of control system or model is to be not smaller, than diversity of the object itself. The *Law of Adequateness*, given by S.Beer, establishes that for optimal control the objects are to be compensated by corresponding black boxes of the control system [1]. For optimal pattern recognition and clustering only partial compensation is necessary. More of what we are interested in is to minimize the degree of compensation by all means to get more accurate results.

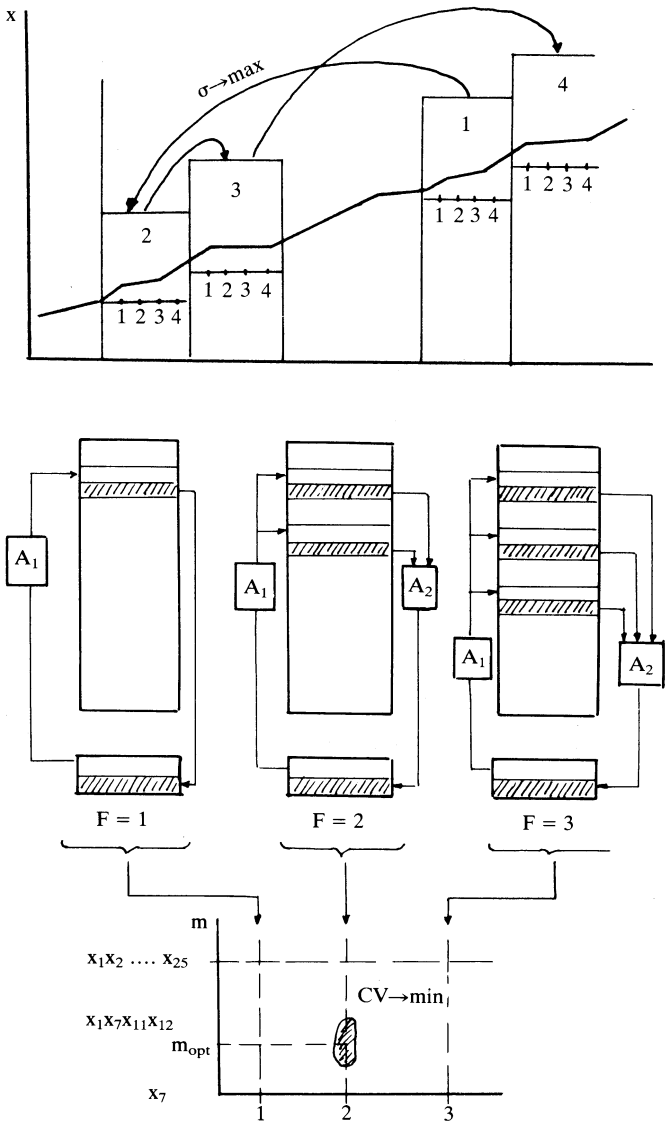

Figure 1

The equal fuzziness of the model and object is reached automatically if the object itself is used for forecasting. This is done by searching analogues from the given data sample which are equivalent to the physical

model. Forecasts are not calculated in the classical sense but selected from the table of observation data.

The main assumptions are the following:

- the system to be modeled is described by a multidimensional process;
- the multidimensional process is sufficiently representative, i.e. the essential system variables are included in the observations and there are enough observations in data sample;
- it is possible that a part of past behavior will be repeated.

This approach is recommended when number of input variables is bigger than number of observations, in opposite case parametric GMDH algorithms may be used [3]. If we succeed in finding for the last part of behavior trajectory (starting pattern), one or more analogous parts in the past (analogous pattern) the prediction can be achieved by applying the known continuation of these analogous patterns (Fig.1).

The point A_1 nearest to each given output point A_0 is called as its first analog. The next after distance point A_2 is called second analog and so on. The point, which follows the first analog in time, A_{1F} is called as first analogs forecast. The point, which follows the second analog in time A2F, is called the second analogs forecast and so on. Forecast is calculated by complexing of optimal number of analogues forecasts.

Using a sliding window which generates the set of possible patterns $\{A_{i,k+1}\}$, where $A_{i,k+1} = (x_i, x_{i+1}, x_{i+2}, ..., x_{i+k})$ and k+1 is the width of sliding window and also of the patterns, the output pattern is $A_k^A = A_{N-k,k+1}$.

The algorithm of selection of the analogous pattern has the following task: for the given output pattern A_k^A it is necessary to select the most similar patterns $A_{i,k+1}, i \in J$ and to evaluate the forecast or extrapolation with the help of these patterns.

Twice-multilayered neuronet is highly increase accuracy of modeling. It is used for optimization of AC algorithm parameters and selection of the optimal set of variables. The full selection task for neuronet is a four-dimensional problem with the following dimensions:

a) set of variables used;
b) number of analogues selected;
c) width of the patterns (number of lines);
d) values of weight coefficients with which patterns are complexed.

But examples showed that big full procedure can be replaced by three simple partial sorting-out procedures. As method of optimization the comparison of variants by value of criterion accuracy is used. The criterion is calculated on the whole length of sample. This is the way of short-time forecasting problem solution on one step ahead. More difficult is the problem of long-term step-by-step random processes forecast. To select similar patterns from all possible patterns in the time series, the following steps are developed:

A. Search of the optimal set of variables

The choice of the optimal set of variables can be realized by sorting of possible sets of variables. In case of small number of variables (<10) full sorting-out of possible variables sets is used and for bigger amount of variables – method of gradually complication of variables sets (reduced sorting) is used.

At first is founded distance *L* from each observation to characteristic points of other observations. Euclidean distance is defined as:

$$L = \sqrt{\sum_{j=1}^{n} \left(x_{ij} - x_{aj}\right)^2} \qquad (1)$$

where x_{ij} - value of *j* variable in *i* analyzed point;

x_{aj} – value of its analogue;

n – number of observations

As analogue is used observation, characteristic point of which is placed more closely to point under investigation.

For all observations in data sample founded its analogues (nearest neighbors). As criterion for optimal variables set selection is used value of accuracy criterion:

$$RR(s) = \sqrt{\frac{\sum_1^N \left(x_i - x_a\right)^2}{\sum_1^N \left(x_i - \bar{x}\right)^2}} \rightarrow \min, \qquad (2)$$

where: x_i - is the variable values in the table;

x_a - is the corresponding analogue value and

$\bar{x}$ is the mean value.

Reducing of the sorting procedure is reached by gradually complication of variables sets based on variable selected at the previous step. If define as N number of investigated variables, than at first sorting is done with N samples. Then N-1 sets of all variables pairs with variable selected at the previous step are used for analogues search.

	Полный перебор признаков	
1	6	0.3194
2	1, 2	0.1687
3	3, 6, 8	0.1915
4	1, 2, 3, 4	0.1751
5	1, 2, 3, 7, 8	0.1795
6	1, 2, 3, 5, 6, 10	0.1691
7	1, 2, 3, 4, 7, 10, 11	0.1466
8	1, 2, 3, 4, 7, 8, 10, 11	0.1471
9	1, 2, 3, 4, 6, 7, 8, 10, 11	**0.1416**
10	1, 2, 3, 4, 5, 6, 8, 9, 10, 11	0.1930
	Ограниченный перебор признаков	
1	6	0.3194
2	3, 6	0.2662
3	3, 6, 8	0.1915
4	3, 6, 8, 11	0.2178
5	3, 6, 8, 10, 11	0.2477
6	1, 3, 6, 8, 10, 11	0.2373
7	1, 2, 3, 6, 8, 10, 11	0.1981
8	1, 2, 3, 6, 7, 8, 10, 11	0.1660
9	1, 2, 3, 4, 6, 7, 8, 10, 11	**0.1416**[a]
10	1, 2, 3, 4, 6, 7, 8, 9, 10, 11	0.1975
11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	0.2014

Table 1

Then tested N-2 sets, which include best variables from previous layers and so on.

Results of computation showed that both types of sorting give practically the same result (table 1). In the table best results for full sorting and reduced sorting procedures of real chemical data are considered. Both procedures are select the same 9 variables, taking out variables x_5 and x_9.

B. Transformation of analogues (not obligatory step)

Most processes in large-scale systems are evolutionary. In this case stationarity as one important condition of successful use of the method of analogues is not fulfilled. As time-series may be non-stationary patterns with similar shapes may have different mean values, standard deviations and trends. In the literature, it is recommended to evaluate the difference between the process and its trend, which is an unknown function of time. Another possibility gives the selection of differences where the criterion of stationarity is used as selection criterion. The results of the method of analogues depend on the selected trend function.

It is advisable to determine transformed patterns $A^*_{i,k+1} = (\underline{x}^*_i, \underline{x}^*_{i+1}, ... \underline{x}^*_{i+k})$, where $x^*_{jt} = w_{j0} + w_{j1} x_{jt}$. The weights w_0, w_1 for each pattern $A_{i,k+1}$, when k>1 can be estimated by means of the least squares method, which gives not only the unknown weights but also the total sum of squares s_i^2 , which can be used in the following step as a measure of similarity.

C. Selection of the most similar analogues

The closest analogue is called the first analogue A_1, the next one in distance A_2 is called the second analogue and so on until the last analogue A_F. Distances can be measured by means of the Euclidean distance of points of the output pattern and the analogue or by other measures of distance. In our case it is not necessary to find a proximity measure, but the total sum of the squares s_i^2 gives us information about the proximity between A_1 and A_F.

D. Combining of forecasts (complexing)

Every selected analogous has its continuation which gives a forecast. In such a way we obtain F forecasts, which are to combine. In the literature there are several principles for combination of forecasts. The unknown predictions x_{N+k}, of the M systems variables can be assumed as a linear combination of the continuations of selected analogous patterns, i.e.:

$$x_{N+j} = g_0 + \sum_{j \in J} g_k x^*_{j+k+i} \ ,$$

The unknown parameters g_0, g_j, $j \in J$, are estimated by means of parametric selection procedures, or using self-organizing algorithms. The only problem may be in the small number of observations, i.e. small number of selected patterns.

2. Self-organization of twice-multilayered neural networks

A neural network is designed to handle a particular task. This may involve relation identification (approximation), pattern and situation recognition, or a forecast of random processes and repetitive events from information contained in a sample of observations over a test or control object.

The present stage of computer technology allows a new approach in neural networks, which increases the accuracy of classical modeling algorithms. Such complex system can solve complex problems. We can use the GMDH algorithms as the complex neurons, where the self-organization processes are well studied. Algorithms are examples of complex active neurons, because they choose the effective inputs and corresponding coefficients of them by themselves, in process of self-organization. The problem of neuronet links structure self-organization is solved in a rather simple way.

Each neuron is an elementary system that handles the same task. The objective sought in combining many neurons into a network is to enhance the accuracy in achieving the assigned task through a better use of input data. As already noted, the function of active neurons can be performed by various recognition systems, notably by Rosenblatt's two-layer perceptrons - such neural network achieves the task of pattern recognition. In the self-organization of a neural network, the exhaustive search is first applied to determine the number of neuron layers and the sets of input and output variables for each neuron. The minimum of the discriminating criterion suggests the variables for which it is advantageous to build a neural network and how many neuron layers should be used. Thus, the theory of neural network self-organization is similar in many respects to that of each active neuron.

Active neurons are able, during the self-organizing process, to estimate which inputs are necessary to minimize the given objective function of the neuron. In the neuronet with such neurons, we shall have twofold multilayered structure: neurons themselves are multilayered, and they will be united into a multilayered network. They can provide generation of new features of special type (the outputs of neurons from previous layer) and the choice of effective set of factors at each layer of neurons. The output variables of previous layers are very effective secondary inputs for the neurons of next layer. First layer of active neurons acts similar to Kalman filter: output set of variables repeated the input set but with filtration of noises. Number of active neurons in each layer is equal to number of variables given in initial data sampling.

Neuronet structure is given in Fig. 2. Solely including the output calculated variables from each previous layer of neurons effects sample extension. The samples show the form of the discrete template used to teach the first neurons of a layer by the Analogues Complexing GMDH algorithm. The algorithm will suggest which of the proposed arguments should be taken into consideration and will help to estimate the connectivity coefficients.

Each layer in a neural network contains neurons, whose outputs correspond each to a particular specified variable: the output of the first neuron to the first variable, the output of the second neuron to the second variable, etc. Each column consists of neurons whose outputs correspond to one of the variables. From each column in turn, one neuron with a minimal variation criterion is selected. More specifically, one neuron having the best result is selected from the first column of neurons for which the output is the first variable; similarly, one neuron is selected from the second column of neurons for which the output is the second variable, etc. This selection procedure uniquely defines the number of layers for each variable and, thus, the structure of the neural network.

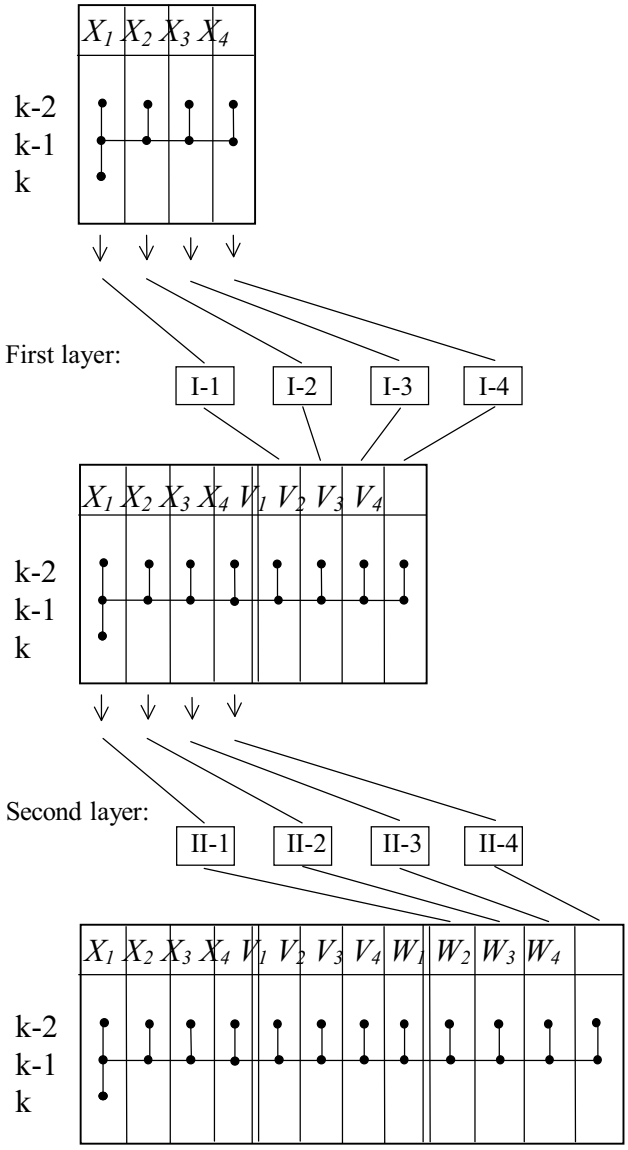

Fig 2. Schematic arrangement of the first two rows of a neural network.

To begin with, we construct the first layer of neurons in the network. Then we will able to determine how accurate the forecast will be for all variables. For this purpose, we use a discrete template that allows a delay of one or two days for all variables. Then we add a second, a third,

etc. layer to the neural network, as shown in Fig. 6, and go on doing so as long as this improves the forecast or decrease external criterion value.

For each neuron, we have applied the extended definition procedure to one model (out of the five closest to the optimal one). For the optimal models, we have calculated the forecast variation RR(s) criterion. It may be inferred, that there is no need to construct a neural network in order to form a forecast for those variables, for which variation criterion value takes on the least value in the first layer. It is advisable to use a neural network to form a forecast for the variables, for which the variation criterion takes on the least value in the last layers of neurons.

The equations for the neurons of the network define the connections that must be implemented in the neural network; in this way they help achieve the task of structural self-organization of the neural network. For brevity, the data sample in the above example is extended in only one way: tile output variables of the first layer are passed on as additional variables to the second, third, etc. layer of neurons. It is possible to compare different schemes of data sample extension by external criterion value.

The task for self-organization of such networks of active neurons by selection is to estimate the number of layers of active neurons and the set of possible potential inputs and outputs of every neuron. The sorting characteristic - ”number of neuronet layers - variables, given in data sample“ - defines the optimum number of layers for each variable separately. Neuronets with active neurons should be applied to raise the accuracy of short-term and long-term forecasts.

Not only GMDH algorithms, but also many modeling or pattern recognition algorithms can be used as active neurons. Its accuracy can be increased in two ways:

- each output of algorithm (active neuron) generate new variable which can be used as a new factor in next layers of neuronet;
- the set of factors can be optimized at each layer. The factors (including new generated) can be ranked after their efficiency and several of the most efficient factors can be used as inputs for next layers of neurons. In usual once-multilayered NN the set of input variables can be chosen once only.

3. Procedure of "divergence and convergence" or "extension and narrowing" of full input data samples

The number of variables in the full sets is extended with each next layer (from two to four in our example). Simultaneously the variables, which are less effective, should be excluded at each layer of neuronet (two variables in example). For this purpose all the variables - candidates should be checked for forecasting ability by analogs complexing and only definite number of the most effective variables (equal to four in example) should be included into full variables set of each next layer of neuronet. In usual neuronet (perceptron) the set of input variables is optimized only one time, at first layer. In twice-multilayered neuronets with active neurons the set of variables can be optimized several times, at all neurons layers input.[4].

4. Example

We apply described above method of search of optimal variables set for extrapolation of chemical activity of two data samples. In first sample for analysis was used series of molecules, analogues of antimicin, which have 'antifilaril' activity.

First input data sample contained 31 compounds and 53 variables. In result of reduced sorting was found set of variables, which contain 10 variables only. At figure 3 shown dependence of RR(s) from steps of sorting. At each step number of considered variables is increased on 1 variable. At last step all variables are tested. At figure is shown that RR_{min}= 0.1699 was founded for 10 variables X={3,13,19,23,24,27,29,32,50,52}. It must be noted that criterion value is raising up during further increasing of variables number.

Second data set had consisted of 35 compounds and 31 variables. The minimal criterion value for defining of molecules activity was equal 0.1237 is received for the following set of variables X={2,7,9,14,16,20,25,26,30}.

Founded variables sets not only he most simple, but the most effective, because for them is reached the minimal average distance between characteristic points of object under investigation, which consequently used all points of sample and their first analogue.

With help of computational experiments was founded that noise dispersion for output variable on one order less than noise dispersion in input data. That means that analogues can be used for noise filtering.

As attachment is given the following results of computation by the current version (15.04.98) of GMDH AC module:

1. R1_KAPPA – results of analisis of compounds activity by three methods of variables selections;
2. R2_KAPPA – results of evaluation of all variables by different criteria;
3. KAPPA – initial data sample, it was normilized before computation.

4.Conclusion

The algorithm of analogues complexing is used instead of mathematical modeling for solution of random processes forecasting problem. This algorithm can be applied for forecasting of all the variables given in input data sample but with different efficiency. Three sorting-out optimization help to increase the accuracy of forecasting. For further increase of accuracy the neural with active neurons can be used where the neurons are the algorithms of analog complexing. The neuronet should be considered as the tool to rise up of processes forecasting accuracy, thanks to the generation of new variables (X_{1F}, X_{2F}, X_{1FF} and so on) and to choice of the most effective variables sets at each layer of neuronet. From biological point of view neuronet with analog active neurons is more similar to natural brain than neuronets using mathematical modeling.

References

1. Beer,S. Cybernetics and Management, English Univ.Press, London, 1959, p.280.
2. Ashby D. An introduction to cybernetics. J. Wiley, New York 1958.
3. Madala,H.R. and Ivakhnenko,A.G. Inductive Learning Algorithms for Complex Systems Modeling. CRC Press Inc., Boca Raton, 1994.
4. Ivakhnenko A.G. An Inductive Sorting Method for the Forecast of Multidimensional Random Processes and Analog Events with the Method of Analog Forecast Complexing. Pattern Recognition and Image Analysis, 1991, Vol.1,No1.pp.101-107.
5. Ivakhnenko A.G. Ivakhnenko G.A. and Mueller J.A. Self-organization of neuronets with Active Neurons. Pattern Recognition and Image Analysis,1994,Vol.4,No.4,pp.177-188.
6. Ivakhnenko A.G. and Mueller J.A. Self-organization of Nets of Active Neurons. System Analysis Modeling Simulation SAMS, 1995, Vol.20,pp.93-106.
7. Ivakhnenko G.A. Self-organization of Neuronet with Active Neurons for Effects of Nuclear Tests Explosions Forecasting. System Analysis Modeling Simulation SAMS, 1995, Vol.20, No.4, pp.107-116.