GMDH Algorithm for Optimal Model Choice by the External Error Criterion with the Extension of Definition by Model Bias and Its Applications to the Committees and Neural Networks

A.G. Ivakhnenko*, E.A. Savchenko*, and G.A. Ivakhnenko**

** International Center of Information Technologies and Systems, pr. Akademika Glushkova 40, Kiev, 03187 Ukraine*
e-mail: gai@gmdh.kiev.ua
http://www.GMDH.net
***National Institute for Strategic Studies, ul. Pirogova 7a, Kiev, 252030 Ukraine*
e-mail: Ivakhnenko@gmdh.net

Abstract—In the case of substantial noise, i.e., for inaccurate and incomplete data, the use of the Group Method of Data Handling (GMDH) algorithm leads to sharp and rather deep minimums of dependency of external criterion of accuracy measured on testing sample on the complexity of model structure. This minimum indicates the optimal model. In practice, however, if the noise is just noticeable, i.e., if data are accurate, the minimum becomes indefinite or does not exist at all. In this case, an extension of definition is needed based on a new criterion such as, e.g., the value of a model bias measured on the two identical data samples. The combinatorial GMDH algorithm with an extension of definition by the model bias can be used as a neuron in committees and in repeatedly multilayered neural networks for solving the problems of medical monitoring.

1. INTRODUCTION

When the data sample is transformed to a conditional form given by Gauss, the various interpolation problems of artificial intelligence, such as pattern recognition, dependence detection, stepwise prediction of random processes, etc., can be solved by general algorithms, mostly by the combinatorial GMDH algorithm [1, 2]. These algorithms differ mostly in the choice of output variables and modeling space coordinates. For pattern recognition and dependence detection, the discriminant functions are found, whereas for prediction, the dependence of the further values of one of the variables on the current and delayed values of all variables is found. The new interpolation-type models [3] accept current and delayed values as coordinates of modeling space. The future values can also be used. A general problem-solving approach means that an algebraic theory used for pattern recognition can be also applied to other problems [4]. Particularly, since there are many ways of generating secondary arguments–candidates, almost always we have a possibility of obtaining a polynomial model linear on coefficients in which the number of members is equal to the number of realizations given in the testing data sample, i.e., to obtain a so-called limiting model with zero error value [5–7].

Thus, in [5], the pattern recognition problem is considered for ten primary and 45 secondary arguments–candidates. The number of realization is 40. This means that C_{55}^{40} limiting models with zero error values can be *a priori* pointed out. It should be noted that such models are not interesting for modeling because they have big bias; i.e., they cannot be generalized to the case of a new data sample. There are two ways of choosing an optimal model with a generalization property. The authors of [5] solve the problem of modeling choice without self-organization, that is, without dividing data sample into two parts. To solve a problem of a model choice, a threshold value of the coefficient of correlation between the given variables and the output variable module is set to 0.3. The less effective variables are excluded from consideration. This gives one a possibility of choosing a model that has no more than 12 arguments from 55 arguments possible. This can be justified by the experience accumulated by the authors during the work with such systems.

A subjective interference of a man can be considered as a first external supplement necessary for the model choice in terms of S. Beer's theory [8]. In the case of self-organization, the input sample should be divided into two or more parts. A human interference is replaced by the direct calculation of the external error criterion of the model on a testing sample. The examples testify that a sharp minimum of dependence is only obtained in the region of incomplete or inaccurate data. However, in the most important in practice area of accurate, complete, and noise-free data, the minimum of dependence becomes indefinite. In this case, an extension of the definition of the optimal model should be used after the bias of the model. The bias criterion

Received May 4, 2002

Pattern Recognition and Image Analysis, Vol. 12, No. 4, 2002, pp. 347–353.

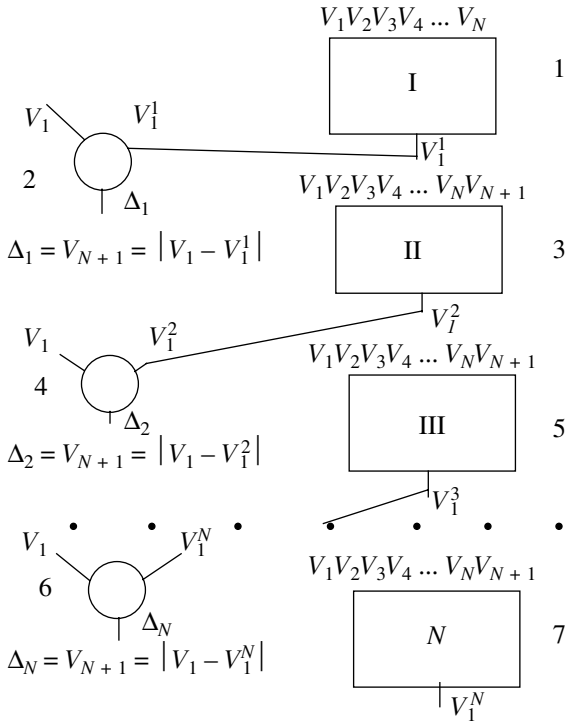

Fig. 1. Iterative algorithm of the twice-multilayered neural net with the difference elements (an example for N input variances); 1, 3, 5, 7 are the neurons working according the MGUA algorithms; 2, 4, 6 are the elements calculating error of modeling module.

can be only calculated on two identical, i.e., equal in size and statistical properties, data samples *A* and *B*. The problem of two-criterion choice of an optimal model does not exist because the properties of accuracy and bias of the model are independent. Bias of the model is equal to zero when the models received on *A* and *B* samples are equal; in this case, however, they can be both inaccurate. That is why an extension of the definition can be proposed. This implies that both criteria should be sequentially used. First, a set of sufficiently accurate models should be chosen and, then, the most unbiased model should be chosen from among them. The condition of model unbiasedness can be considered as an objective second external supplement [8]. As was noted above, if the data are accurate and complete, the minimum of the external criterion is not well defined and sometimes does not exist. The sorting out procedure is performed in such a way that the use of new criterion, e.g., the model bias, is needed. We can conclude that by dividing the data sample into two parts *A* and *B*, we give new properties to the sorting-out process. A solution to the problem of multialternative pattern recognition needs the data sample to be divided into several parts. This is used for choosing an optimal project by constricting a committee of pattern recognition systems. The committee is trained to evaluate the project in the same way as a team of experts does. The GMDH algorithm with additional definition of optimal model by its bias is used as a neuron in the committees and repeatedly multilayered neural networks. In the neural networks, each neuron receives information from all of the neurons of the preceding layers. In the committee, each neuron receives its special independent part of information about the object.

It is possible to construct neural networks where a committee of neurons realizes each neuron. A neural network can be considered as a sequential combination of Kalman-type filters [9]. In this case, a number of layers for each variable can be different depending on the noise variance added to each variable. From another point of view, a neural network can be considered as a convenient way of generating different secondary arguments–candidates. In Fig. 1, the twice-multilayered neural network is shown. This network has different elements, which calculate the model error in each layer. The error is sent to the next layers of neurons as one of the effective arguments–candidates. The number of the layers is growing up until the external error criterion is decreasing. In the neural networks, an extension of definition of the optimal number of layers is not necessary, because the number of neurons in each layer is constantly equal to the number of variables in data sample. An additional definition is only necessary in the case when the number of variables is modified. It is convenient to perform different transformations of variables necessary for generating the secondary arguments–candidates. Particularly, it is possible to calculate integral characteristics of chaotic processes necessary for predicting the processes in such objects. Finally, we consider the problem of developing through a learning process the computer language, which corresponds to the language of an expert team. It is necessary to transform a non-formalized natural language to a formalized language of computer, of course, with a certain increase in the number of elements for language codes.

2. SELF-ORGANIZATION OF DISCRIMINANT MODEL FOR RECOGNIZING ARCHAEOLOGICAL FINDINGS WITH EXTENSION OF A DEFINITION BY MODEL BIAS

Pattern recognition of archaeological findings is described in detail in [5]. Different physical properties of findings are used as the primary arguments (ten in total). The pairwise covariations of the primary arguments are used as the secondary arguments. For brevity, we shall not take the secondary arguments into account although they certainly increase an accuracy of a model. An input data sample was ranked according to dispersion of variables and, then, divided into two samples *A* and *B* (see table). As table shows, all variables are binarized, i.e., replaced by values +1 or –1. The binarization of variables is associated with noise added both to arguments and output variables. Moreover, the level of noise variance can be considered typical of practical problems. Thus, if a combinatorial algorithm

Table 1. Initial data samples **A** and **B**

A

Serial number	Number in the initial data sample	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y
1	2	–1	1	1	–1	–1	–1	1	1	–1	1	1
2	3	1	1	1	–1	–1	–1	1	1	–1	1	1
3	6	–1	1	–1	1	–1	–1	1	1	1	–1	1
4	7	1	1	–1	–1	–1	–1	–1	1	–1	1	1
5	10	–1	–1	–1	–1	–1	1	1	–1	1	1	1
6	13	–1	1	1	–1	–1	1	1	1	1	1	1
7	14	1	1	1	–1	–1	–1	1	1	1	–1	1
8	15	1	1	1	–1	–1	1	1	–1	–1	–1	1
9	19	1	1	1	–1	–1	1	1	1	1	1	1
10	20	–1	1	–1	–1	1	–1	1	–1	–1	–1	1
11	24	1	1	–1	1	1	1	–1	–1	1	–1	1
12	26	1	1	1	–1	–1	1	–1	1	1	1	1
13	28	–1	1	–1	–1	–1	–1	1	–1	1	1	–1
14	30	–1	1	–1	–1	–1	–1	1	1	1	1	–1
15	31	–1	1	–1	–1	1	–1	1	1	1	1	–1
16	34	–1	1	1	1	1	–1	–1	–1	1	1	–1
17	35	–1	–1	1	–1	–1	–1	1	–1	1	–1	–1
18	37	–1	–1	–1	–1	1	–1	1	–1	1	1	–1
19	39	1	1	–1	–1	–1	–1	1	–1	1	1	–1
20	40	–1	1	–1	1	1	–1	–1	–1	1	–1	–1

B

Serial number	Number in the initial data sample	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y
1	1	–1	1	1	–1	1	–1	1	1	1	–1	1
2	4	–1	1	1	1	–1	–1	–1	1	1	–1	1
3	5	1	1	1	–1	1	–1	1	–1	–1	1	1
4	8	1	1	1	1	1	–1	–1	–1	1	–1	1
5	9	–1	1	1	–1	–1	–1	1	1	1	–1	1
6	11	1	1	1	–1	–1	1	–1	1	–1	1	1
7	12	–1	1	1	–1	–1	–1	1	–1	–1	–1	1
8	16	1	1	–1	–1	–1	1	1	1	1	1	1
9	17	–1	1	1	1	–1	1	–1	1	1	–1	1
10	18	–1	–1	1	–1	–1	–1	–1	–1	–1	1	1
11	21	1	1	1	1	1	1	–1	1	1	1	1
12	22	1	1	1	–1	–1	–1	–1	1	–1	1	1
13	23	–1	–1	–1	–1	–1	–1	1	1	–1	1	1
14	25	–1	1	1	–1	–1	–1	1	–1	–1	1	1
15	27	1	1	–1	1	–1	1	1	1	1	–1	1
16	29	–1	1	–1	–1	1	1	1	–1	1	1	–1
17	32	–1	1	–1	–1	1	–1	1	–1	1	1	–1
18	33	1	1	–1	1	1	–1	1	–1	1	–1	–1
19	36	1	1	1	–1	1	–1	1	–1	1	–1	–1
20	38	1	1	–1	–1	1	–1	1	–1	1	1	–1

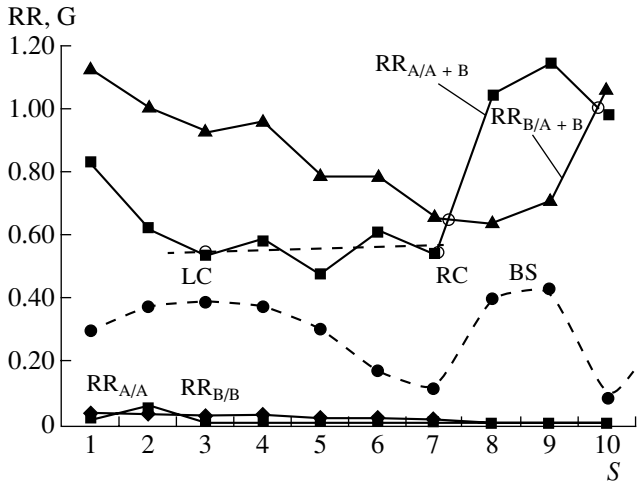

Fig. 2. Accuracy criterion and biases against the model complexity.

with additional definition of bias solves the given problem, we can conclude that it has enough noise immunity for solving most of the similar problems.

The dependence of criteria of accuracy and bias on the complexity of a model is shown in Fig. 2. A set of models–candidates includes ten models made up for the most effective arguments, which have the bigger values of the correlation coefficient module. Calculations show that the external error criterion $RR_{A/A+B}$, first, quickly goes down, then, oscillates around some middle value, and, finally, goes up. In Fig. 2, *LC* and *RC* are the left and the right corners of a characteristic, respectively. A dotted line denotes a middle value of the external error criterion $RR_{A/A+B}$, a center of insignificant oscillations of values. All models beginning from *LC* to *RC* inclusive are subjected to the additional definition procedure according to the bias criterion.

The algorithm for using the bias criterion can be described as follows. All the models that generate the indefinite minimum of external error criterion calculated on the testing sample are considered to be models–candidates and are evaluated according to the criterion of model bias. For each of them, the estimates of a coefficient are calculated by mean-square error on samples *A* and *B*, and this gives us a possibility of calculating error criteria $RR_{A/A+B}$ and $RR_{B/A+B}$.

Further, for the six models chosen, the errors are calculated on the sample *A*, if the estimates of coefficients were calculated on the sample *B*, and on the sample *B* if the estimates of coefficients were calculated on the sample *A*.

The model bias is calculated by formula: $BS = |RR_{A/A+B} - RR_{B/A+B}| \longrightarrow \min$.

Thus, we obtain the following optimal model:

$$y = 0.7124 + 0.1209x_1 - 0.1274x_2 - 0.2171x_3 - 0.1693x_5 + 0.3374x_8 - 0.4382x_9.$$

The model accuracy is characterized by the following values of criteria: MSE = 0.370; MAPE = 49.32%, where MSE is a mean-square error and MAPE is a mean absolute percentage error of modeling. A minimum of the bias $BS_{\min} = 0.013$ indicates the optimal model, which has the maximum of generalization properties.

Figure 2 also shows two polynomial discriminant models with seven and ten most effective arguments, respectively, which have closest to zero bias *BS*; i.e., they have the maximum of generalization properties. Here, the model with seven arguments should be chosen, since it has the lesser bias for the lower error value.

3. A CHOICE OF MODELING SPACE COORDINATES

For computing time to be reasonable, about 20 most effective arguments should be chosen from the set of all primary and secondary arguments–candidates to be the coordinates of a modeling space. The optimal set of coordinates can be chosen by a module of correlation coefficient or by filtration of effective variables described above.

3.1. Ranking Arguments–Candidates According to the Correlation Coefficient of the Argument with the Output Variable Module

In the example of recognition of archaeological findings, the most effective are the following 18 arguments: x_6x_9, x_2x_3, x_5, x_8x_9, x_4x_9, x_8, x_3, x_2x_8, x_9x_{10}, x_2x_5, x_3x_4, x_4x_5, x_2, x_3x_{10}, x_1x_{10}, x_7x_9, x_9, x_5x_{10}.

Using the combinatorial GMDH algorithm, we find the following model that contains 12 members:

$$y = 0.0088 + 0.269x_2 - 0.2866x_3 - 0.3899x_5 - 0.165x_9 + 0.505x_2x_3 + 0.265x_4x_9 + 0.382x_2x_3 + 0.323x_2x_3 + 0.235x_2x_8 - 0.105x_9x_{10} - 0.138x_5x_{10}.$$

A model accuracy is characterized by the following values of error criteria: MCC = 0.865; CR = 0.267; MSE = 0.469; and MAPE = 34.11%, where MCC is a correlation coefficient of the model and CR is an external criterion.

3.2. Multiple Use of Combinatorial GMDH Algorithm

To apply a filtration, the following set of 55 arguments–candidates ranked over the module of coefficient correlation is split into three separate subsets of arguments:

(1) 20 most effective, according to the module of correlation coefficient, arguments: x_5, x_2x_3, x_6x_9, x_8x_9, x_4x_9, x_8, x_3, x_2x_8, x_9x_{10}, x_2x_5, x_3x_4, x_4x_5, x_2, x_3x_{10}, x_1x_{10}, x_7x_9, x_9, x_5x_{10}, x_2x_{10}, x_5x_8.

(2) 20 less effective arguments: x_5x_7, x_3x_6, x_1x_6, x_1x_2, x_8x_{10}, x_7x_{10}, x_6x_{10}, x_4x_{10}, x_3x_9, x_1x_9, x_7x_8, x_4x_8, x_3x_7, x_1x_7, x_4, x_1, x_5x_9, x_2x_9, x_6x_8, x_6x_7.

(3) 15 least effective arguments: x_3x_5, x_1x_5, x_2x_4, x_6, x_3x_8, x_1x_8, x_4x_7, x_5x_6, x_2x_6, x_1x_4, x_1x_3, x_{10}, x_7, x_2x_7, x_4x_6.

For each set, by using the combinatorial GMDH algorithm, the model can be found and the error criterion can be calculated. All the arguments included in any set of effective variables comprise a set of 18 effective arguments selected by the combinatorial GMDH algorithm: x_2x_3, x_4x_9, x_2x_8, x_4x_5, x_6x_9, x_9x_{10}, x_1x_{10}, x_3x_{10}, x_1x_6, x_2x_{10}, x_7x_{10}, x_6x_{10}, x_4x_{10}, x_3x_9, x_7x_8, x_4x_8, x_1x_7, x_2x_9.

By using the combinatorial GMDH algorithm we find the following model containing 13 members:

$$x = 0.1397 + 0.203x_2x_3 + 0.258x_4x_9 + 0.213x_2x_8 + 0.180x_4x_5 + 0.381x_6x_9 - 0.127x_9x_{10} + 0.084x_1x_{10} - 0.048x_3x_{10} + 0.262x_2x_{10} - 0.207x_7x_{10} + 0.054x_3x_9 + 0.002x_7x_8 + 0.22x_2x_9.$$

A model accuracy is characterized by following values of error criteria: MCC = 0.906; CR = 0.396; MSE = 0.397; and MAPE = 31.75%.

We can conclude that filtration proves to yield more accurate results here and, moreover, the obtained model includes no primary features, but only their pair covariance.

4. USING THE GMDH ALGORITHM FOR FILTERING A DATA SAMPLE

A similarity between the combinatorial GMDH algorithm [1, 2] and the Kalman filter [9] consists in the fact that they both use a simplified model of an object. The difference is that, in the Kalman filter, a model is given by a human, an author of filtration, whereas in the GMDH algorithm, it is obtained by a self-organization, i.e., by sorting out on the two data samples. The GMDH filter can be used for the solving different problems. Let us point out two of them.

(1) A choice of an optimal model when the noise dispersion is small. It was shown that a sequential sorting of the models according to the external error criterion and to the bias criterion could be considered as a preliminary noise filtration of variables which are then sent to the second sorting out procedure according to the bias criterion.

(2) A second example is the use of the GMDH algorithms for the definition of the modeling coordinate set. Out of a total of 55 primary and secondary arguments a set of 20 coordinates should be chosen which are convenient for application of the combinatorial GMDH algorithm. The problem is to point out this set of 20 variables.

5. A COMMITTEE OF PATTERN RECOGNITION SYSTEMS DIVIDING DATA SAMPLE INTO THREE PARTS

It is of interest to consider why cannot we divide a sample into three parts and use one or another optimization several times. It is possible to develop and investigate an algorithm where not one but several models are chosen from a set of most accurate models on a learning sample. Then, on the next test sample, the optimal model is selected. It is interesting to clarify the potentials of ding a sample into three or four parts. Division of a data sample into several parts is used for constructing the committees of pattern recognition systems and the repeatedly multilayered neural networks.

Let us give an example where three or more independent samples are used. In [10], the development of algorithms for choosing the optimal project for reconstructing Venceslas Square in Prague is described on the basis of proceeding of expert estimations. The experts were divided into learning and testing teams. The actions of two teams were similar to the actions of two data samples in the combinatorial GMDH algorithm. The first team (analogous to the learning sample) consequently generates the more and more complicated projects and, by sorting out variants, optimizes the project parameters according to its own conception (a criterion). Such parameters as avenue widths, heights of buildings, arrangement of plants, etc., are optimized. After that, the project can be subjectively chosen.

However, to increase optimality and objectivity of a choice, the projects are estimated on a three-point scale: excellent, good, and satisfactory. Projects with high estimates are sent to the second team (analogous to the testing sample) which should choose an optimal project. The result of the first teamwork can be represented as the three separate samples which connect parameters of the project with their estimates. By using the samples thus received it is possible to construct a committee of three pattern recognition systems to continue the work of experts. The aim of the committee is to continue the work of the first group of experts after they finish their work. Similarly, the next pattern recognition system can be trained to continue the work of the second teamwork. As a result, we shall receive a two-layered committee which, after learning, can replace both expert teams. On the whole, this system is designed for estimating the projects, which arrive after the experts have finished their work, in the same way as the experts do. A similar system is developed for estimating the global projects of investments of capital in foreign countries [11].

6. REPEATEDLY MULTILAYERED NEURAL NETWORKS AND COMMITTEES OF PATTERN RECOGNITION SYSTEMS USING GMDH ALGORITHMS

It is obvious that designing an artificial intelligence, a human cannot overcome the development of a nature brain because it has an astronomical number of neurons (up to10^7 elements). Therefore, it is inevitable to use more complicated neurons, for example, blocks acting according to the GMDH algorithms. The main difference between the GMDH neurons and the perceptrons is that, in the perceptron, the hidden elements of a sec-

ond layer are not divided into clusters [12] during training. For example, in the pattern recognition system described above, two clusters are indicated: one cluster with the output variable +1 and second cluster with the output variable –1. A new input signal is referred to one or another cluster according to the output of a discriminant model with the subsequent rounding off to +1 or –1. In perceptrons, where there is no such a division of hidden elements, the information about the cluster on which the biggest signal is received is lost. The GMDH neurons can be united into repeatedly multilayered networks and pattern recognition committees to increase the efficiency of solution of the artificial intelligence problems (accuracy and unbiasedness).

7. PROBLEMS OF PATTERN RECOGNITION AND PREDICTION OF RANDOM PROCESSES FOR CHAOTIC OBJECTS

Usually, by chaotic objects we mean the objects which have many random links with equally probable outputs [13]. Previously, at the beginning of investigations, chaotic systems were called indeterministic systems [14]. Classical example is a toss of the coin. To predict an output (heads or tails) is principally impossible. Single events in chaotic objects cannot be predicted.

Nevertheless, the experiments show that the integral characteristics of chaotic processes, which are received by averaging the variables by intervals, can be predicted. Average values can be predicted as usual processes in the presence of random noise. To predict them, the combinatorial GMDH algorithm with additional definition can be applied. The more the length-averaging interval, the more effective predictions. As a basis for such a conclusion we can cite an example of 10000 tosses of the coin; the result is published in [15]. The experiment shows that the probability of the head outcome equal to one half is very rarely observed and, moreover, it continues to change by the arcsine law, i.e., further it is observed more and more rarely. Regarding the twice-multilayered neural networks with GMDH neurons, we can say that there is a definite freedom of choice of different arguments and features–candidates, including obtaining integral characteristic of random processes, i.e., the average on interval variables. For example, in the twice-multilayered neural network shown in Fig.1, the averaging block can be used instead of the difference elements for averaging some randomly changing chaotic variable.

The authors encountered a chaotic object in the problem of predicting time of the next epileptic seizure from the encephalogram signals. Previously averaged characteristics allowed us to find the most effective electrode from among 32 electrodes inserted in the brain. By averaging the voltage of the most informative electrode we can calculate the position of a disease centre and use a neural network with GMDH neurons to predict a time of next seizure. The sufficient time of averaging electrode voltage deflection was one second. The lead-time of seizure prediction was about one minute.

8. SUBSTITUTION OF NON-FORMALIZABLE NATURAL LANGUAGE BY THE LANGUAGE OF CODES CONVENIENT FOR COMPUTERS

The formal computer languages often cannot describe the classes of vector signals easily recognized by a human. Such a situation we encounter, for example, in the problem of recognizing interconnection of the neurons of a brain [16]. An experienced physiologist easily attributes the correlation histograms to one of the ten classes but he or she cannot explain how he or she does it. In order that computers, after learning, continue the work of physiologists, a committee of ten pattern recognition systems was programmed to relate new input signals to one of the ten classes, which are recognized by a human. Pattern recognition systems worked with 27 easily formalizable binary features for each histogram. In this example, a committee of ten pattern recognition systems was used for deciphering a non-formalizable language of a human–physiologist [17]. The input data sample for each recognition system contains a cluster of observations of a given class and a cluster of the other classes of observation. Only ten of all typical realizations of each class were used; on the whole, 90 realizations.

REFERENCES

1. Ivakhnenko, A.G. and Stepashko, V.S., *Pomekhoustoichivost' modelirovaniya* (Noise Immunity of Modeling), Kiev: Naukova Dumka, 1985.
2. Madala, H.R. and Ivakhnenko, A.G., *Inductive Learning Algorithms for Complex Systems Modeling*, Boca Raton: CRC, 1994.
3. Ivakhnenko, A.G., Savchenko, E.A., Ivakhnenko, G.A., and Gergei, T., Using GMDH Algorithms for Reconstruction of Missing Data and Prediction of the Blood Glucose Level During Self-Monitoring of Diabetes, submitted for publication in *Problems of Control and Informatics.*
4. Zhuravlev, Yu.I. and Gurevich, I.B., Pattern Recognition and Image Recognition, *Pattern Recognit. Image Anal.*, 1991, vol. 1, no. 2, pp. 149 – 181.
5. Krug, G.M. and Krug, O.Yu., A Mathematical Method for Classifying Ancient Pottery, in *Tr. Inst. Arkheol. Akad. Nauk SSSR*, Moscow: Nauka, 1965, pp. 317–323.
6. Ivakhnenko, A.G., Kovalishin, V.V., Tetko, I.V., Luik, A.I., and Ivakhnenko, G.A., Self-Organization of Neural Networks with Active Neurons for Prediction of Activity of the Chemical Compounds on the Basis of the Analog Searching Algorithm, *Problems of Control and Informatics*, 1999, no. 1, pp. 69—77.

7. Ivakhnenko, A.G., Ivakhnenko, G.A., Savchenko, E.A., and Gergei, T., Twice Multilayered Neural Network Self-Organization for Interpolation Problems of Artificial Intelligence Solution in the Case of Complete Absence of Information about Small Part of Input Variables, *Proc. 15 Int. Conf. on Modeling*, Prague, 2001.
8. Beer, S., *Cybernetics and Management*, London: English Univ., 1959.
9. Kalman, R.E. and Bucy, R., New Result in Linear Filtering and Prediction Theory, *J. Basic Eng.*, 1961, vol. 83, pp. 95–108.
10. A New Vision for the Heart of Prague: The Planning Weekend for Venceslas Square, *The Prince of Wales's Projects Office*, Lucerna Palace, Prague, 1996.
11. Ivakhnenko, A.G., Savchenko E.A., Ivakhnenko, G.A., and Gergei, T., The Use of the Multialternative Recognition Algorithms and the Method of Group Data Handling for Processing Expertises of the Projects of Global Investments, *Cybernetics and Computer Science*, 2001, no. 133.
12. Ivakhnenko, A.G., Ivakhnenko, G.A., Savchenko E.A., and Wunsch, D., Algebraic Approach and Optimal Physical Clusterization in Interpolation Problems of Artificial Intelligence, *Pattern Recognit. Image Anal.*, 2000, vol. 10, no. 3, pp.404–409.
13. Iasemidis, L.D. and Sackellares, J.C., Chaos Theory and Epilepsy, in *The Neuroscientist*, Williams and Wilkins, 1996, pp. 118–208.
14. Ivakhnenko, A.G. *et al.*, *Prinyatie reshenii na osnove samoorganizatsii* (Decision Making on the Basis of Self-Organization), Moscow: Sovetskoe Radio, 1976.
15. Feller M., *Probability Theory*, Moscow: Nauka, 1985.
16. Villa, A.E.P., Tetko, I.V., Ivakhnenko, A.G., Ivakhnenko, G.A., and Sarychev, A.P. Recognition of the Type of Interaction between Neurons from Their Cross-Correlation Histograms with the Use of the Voting Procedure, *Pattern Recognit. Image Anal.*, 2001, vol. 11, no. 4, pp. 743–750.
17. O'Connor, J. and Seymour, J. *Introducing Neuro-Linguistic Programming*, Thorsons, 1995.

Aleksei G. Ivakhnenko. Born 1913. Graduated from the Leningrad Institute of Electrical Engineering in 1938. Received his PhD (Kandidat Nauk) Degree in Electrical Sciences in 1943 and Doctoral (Doktor Nauk) degree in 1953. Author of the Group Method of Data Handling (GMDH), which is widely used in modeling. Director's Advisor at the Glushkov Institute of Cybernetics, National Academy of Sciences of Ukraine. Corresponding Member of the National Academy of Sciences of Ukraine. Scientific interests: inductive sorting modeling methods for forecasting random processes in fuzzy systems of ecology, biology, medicine, and economics. Author of 15 monographs and more than 300 papers.

Grigorii A. Ivakhnenko. Born 1966. Graduated from the Kiev Polytechnic Institute in 1989. Leading Specialist in the National Institute for Strategic Studies. Scientific interests: data mining and complex systems analysis by inductive methods, pattern recognition and clusterization. Author of 25 papers.

Evgeniya A. Savchenko. Born 1969. Graduated from the Kiev Polytechnical Institute in 1992. In 1995–1998, worked as an engineer in the Institute of Complex Transport Problems. Now, a postgraduate at the International Center of Information Technologies and Systems of National Academy of Sciences of Ukraine. Scientific interests: complex objects modeling, pattern recognition, and forecasting random processes by the GMDH approach.

SPELL: combinatorial, unbiasedness, multialternative, covariations, binarized, binarization, perceptrons, perceptron, arcsine, centre, NON-FORMALIZABLE, formalizable, clusterization