Problems of Further Development of GMDH Algorithms: Part 2

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

**Inductive Systems Department, International Center of Information Technologies and Systems, pr. Akademika Glushkova 40, Kiev, 03170 Ukraine*
e-mail: gai@gmdh.kiev.ua
http://www.GMDH.net

***National Institute for Strategic Studies, ul. Pirogova 7a, Kiev, 252030 Ukraine*
e-mail: gai@gmdh.net
****University of Missouri-Rolla, Missouri, USA*

Abstract—To increase accuracy in interpolation problems of artificial intelligence (pattern recognition, dependence detection, object identification, stepwise forecasting of random processes, etc.), the inductive algorithms are integrated in more extensive algorithms which consist of several gradually complicated stages of searching for the output value optimization. At the first stage, a simple threshold analysis of the efficiency of the input variables is performed. At the final stage, the twofold and threefold multirow neural networks with active neurons are self-organized. Geometrically, the steps of complication can be represented as a gradient descent along the axis of accuracy. Each step of descent should increase the accuracy of problem solving, which is controlled by the depth of the minimum of the external accuracy criterion. At each step of descent, the main problems that need to be further developed and investigated are considered. Beginning with the second step, each step of descent can include both the search GMDH algorithms and the search for the analogs from the history according to some external criterion.

1. THE CONCEPT OF OPTIMIZATION INDEPENDENCE

Not only the search for model structures or analogs, generated according to a predefined reference function, but also a number of gradually complicated algorithms for solving interpolation problems of artificial intelligence (based on such a search) are to be further developed and investigated.

The complication of the GMDH algorithm and of its optimization undergoes four stages or steps of descent. At the first step of descent, the common multirow GMDH algorithms are used. Here, we find an optimal physical clustering and choose necessary clusters; in addition, we exclude the points (sample rows) which are not efficient enough. At the same time, the variable space is extended due to generation of the so-called secondary features. As a result of the first step of descent, many efficient input variables are found, which are fed in the GMDH algorithms. Clustering decreases the number of rows in the data sample, and modeling (performed on the second step of descent) reduces the number of columns of the initial sample to the required minimum. These two independent operations are performed in turn. At the second step of descent, perceptron-type algorithms are used [1]. The third step of descent is the only step where the number of rows and columns are decreased simultaneously, by gradually increasing the efficiency threshold for rows and columns. To do this, an extended combinatorial GMDH algorithm is used at the third step of descent. At the fourth step of descent, the twofold and threefold multirow neural networks are constructed by using one of the above-mentioned combinatorial GMDH algorithms as active neurons.

Usually, the more complex the GMDH algorithm, the more accurate the solution to the problem. In practice, however, one should confine oneself to a certain degree of the algorithm's complexity. This is due to the following reasons:

—if the data are corrupted by noise, increasing complexity of the GMDH algorithm may not yield the more accurate solution;

—the more complex GMDH algorithms may require more information than that contained in the data sample.

To choose an optimal degree of the algorithm's complexity, one can use the *concept of optimization independence*. According to this concept, all the subsequent optimizations do not change the preceding ones, and such independence makes it possible to refer to optimizations as the stages or steps of descent along the axis of accuracy in solving interpolation problems.

There are two ways of performing all four steps of descent.

(a) by using the model's self-organization;

Received July 28, 2000

Pattern Recognition and Image Analysis, Vol. 12, No. 1, 2002, pp. 6–18.

(b) by using model-free algorithms, i.e., by searching for the analogs in the history of the investigated object, when the object itself is used as a model [3, 7].

Figure 1 illustrates the concept of the four steps of descent. The first step begins with the coefficient of determination $Q = 0$ and $RR = 1$.

If the initial data are accurate and complete, then the rule "the more complex the algorithm, the more accurate it is" is followed. However, if the data are imperfect and noisy, a so-called overfitting of the algorithm can occur. In this case, a transition from less complex algorithms to more complex ones does not improve accuracy in solving interpolation problems. If the self-organization is used (i.e., the variants are searched for according to the external criterion), the algorithm's complexity which yields the maximum accuracy is set automatically. For example, working with the twofold multirow neural network may reveal that it should consist of one row of active neurons; i.e., there is no need in using it. Next, it may also happen that, when the threefold multirow neural network is self-organized, better results are obtained if all variables are predicted from the same data sample; i.e., the process of the algorithm's complication can be halted on the simpler twofold multirow neural network, etc. Thus, self-organizing algorithms can achieve the optimal complexity.

We can state a certain similarity of the choice of the optimal complexity of the model structure according to some external criterion and the choice of the degree of complexity of the GMDH algorithm during its descent along the complexity axis. Both problems can be regarded as examples of problems of the optimal complexity choice of the device for receiving inaccurate and noisy data, where the Shannon paradigm is valid [4]. It is known that, according to Shannon's Second Theorem, the complexity of the optimal receiving device decreases with increasing noise variance [5].

In both problems, the risk of overfitting arises only if the initial data are inaccurate or noisy.

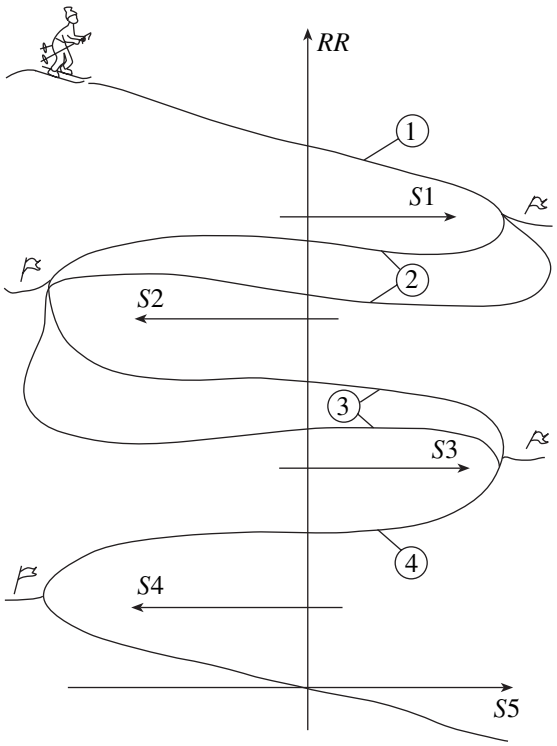


Fig. 1. Picture explaining the concept of optimization independence: 1—a descent by using the common GMDH algorithms; 2—a descent by using the perceptron-type algorithms; 3—a descent by using the extended combinatorial GMDH algorithm; 4—a descent by constructing neural networks with the active neurons; *S*1–*S*5—the complexity of the algorithm; *RR*—an external criterion.

2. FIRST STEP OF DESCENT: GENERATION OF THE SECONDARY FEATURES OR ARGUMENTS–CANDIDATES, THEIR BEING ESTIMATED AND CHOSEN ACCORDING TO SOME EFFICIENCY THRESHOLD

The main problem at the first descent step consists in the choice of the efficient data sample to be subsequently processed. The important role here belongs to the efficiency analysis of the primary and secondary input variables. The primary features or model arguments are the variables taken from the initial data sample. The secondary features are the variables whose values are calculated by using simple functions of the primary features. Most often the covariance of the primary features (the products of normalized values [6]), the coordinates of the first analogs [3, 7], and the output estimates of the variables obtained by GMDH algorithms [2] are used.

All the secondary features–arguments are ranked according to their efficiency regardless of their origin. The efficiency threshold is chosen and the features-arguments whose efficiencies are below this threshold are excluded from the consideration [6]. Only efficient primary and secondary input variables are inputted in GMDH algorithm. For the continuous input variables, the modulus of correlation coefficient estimated by the variable with the output value is usually regarded as the efficiency criterion. For the binary features, such criterion may be the number of resolved arguments [8].

3. SECOND STEP OF DESCENT: THE CHOICE OF REFERENCE FUNCTION AND OF THE WAY OF SEPARATING INITIAL DATA INTO LEARNING AND TEST SAMPLES

The main problems of the second step of descent consist in the choice of the reference function and of the way of separating initial data into the compact sets of the points or clusters which correspond to the classes (patterns) under recognition or to the levels of the predicted process. The reference function is necessary for

generating the set of models–candidates and for estimating them according to the external criterion. Until now, a very limited number of different reference functions are employed. Most researchers use a linear, with respect to the coefficients, polynomial as the reference functions [9]. They justify it by the fact that this polynomial is a discrete analog of the general function decomposition into the Volterra series, although only a small part in the beginning of this series is actually used. Several examples prove that, for most modeling tasks, the more efficient models are obtained for the fractionally polynomial reference functions considered in [10]. A separate line of modeling uses harmonic and exponential reference functions, but only the cases with the small number of time variables are considered [9]. The problem of choosing reference functions is still to be considered.

Another insufficiently developed problem consists in separating data into learning and test parts in order to obtain the external criterion of the model evaluation. It is stated in [12] that an optimal (according to any criterion) model can be obtained only by using information from without, which is called the external complement. This complement either is introduced by an expert (who performs the modeling) or it takes the form of the external criterion, which is calculated on a separate test sample of the initial data. In practice, two main ways of separating initial sample are worth considering.

The first one uses the so-called cross-validation criterion. The test part consists of all sample rows taken in turn, i.e., all cluster points in turn, and the results are averaged. The cross-validation criterion is time-consuming and, therefore, it can be applied only to samples with no more than a hundred rows. Another external criterion is called the regularity criterion. To calculate it, all points of the cluster reduced to the single-moment form are ranked according to the variance and each third point goes to the test part. The rest of the points are used to obtain the coefficient estimates by the Gaussian method or the Khaletskii procedure. If the number of arguments is too high, the size of the test sample can be reduced. For the polynomial models, one can use the formula for changing the rate of data selection to the test sample: $\Delta = \sqrt{2S}$, where S is a number of polynomial terms. If, for example, there are two terms in the polynomial, the test sample should consist of each second point of the sample ranked according to variance.

In the problems of pattern recognition, the input vector is assigned to a cluster according to the proximity criterion. A single cluster division is sufficient for the combinatorial GMDH algorithm. In the iterative or adaptive algorithms, the external character of the accuracy criterion is reduced with each iteration. This manifests itself in less sharp minimums of the criterion. Thus, the problem of pertaining the external character by the criterion through iterations arises. Perhaps, separating the cluster's points should be repeated before each iteration. The advantage of the combinatorial algorithm becomes evident.

3.1. Two Forms of the External Accuracy Criterion.

In the problems of step-by-step prognosis and dependence detection, it is convenient to use the relative form of external criterion [10]:

$$\mathbf{RR} = \sqrt{\sum_{n=A+1}^{A+B}(x_i - \hat{x}_i)^2 \Big/ \sum_{n=A+1}^{A+B}(x_i - \bar{x}_i)^2} \longrightarrow \min,$$

where x_i is a real value of the variable, $\hat{x}_i$ is its value found from a model, $\bar{x}_i$ is an average value of the predicted variable, A is the number of rows in a learning sample, and B is the number of rows in a test sample.

The averaging interval should be the same for all models and algorithms compared. The external criterion is calculated on the independent material, i.e., on the separate data sample which contains N rows. The determination criterion $Q = \sqrt{1 - RR^2} \longrightarrow \max$ could be practically equivalent.

In pattern recognition problems, the determination of an average value is connected with some difficulties. It should be *a priori* known and equal to the most probable value of the discriminant function. Here, the accuracy is determined by the number of recognition errors on the independent material [6]. The accuracy of a recognition model can also be determined by the proximity of the determinant function to the value +1, which corresponds to the accurate recognition. In this case, we can calculate the absolute criterion of regularity [5],

$$AB = \sqrt{\frac{1}{N}\sum_{i=A+1}^{A+B}(1 - \hat{x}_i)^2},$$

if the total number of rows is $N = A + B$.

4. THIRD STEP OF DESCENT: ADDITIONAL OPERATIONS IN A COMBINATORIAL GMDH ALGORITHM IN THE CASE WHEN A PROCESSED CLUSTER CONTAINS TOO MANY (OR FEW) POINTS

To estimate model coefficients by using the least square method (the Gaussian procedure), the number of sample rows should be two or three times as many as the number of terms in the polynomial model. In many clusters subject to processing, this requirement is not fulfilled. In this case, we recommend that the initial data sample should be extended by taking into account the coordinates of the arithmetic and geometrical mean points according to the following formulas:

—with the geometrical mean coordinates $\mathbf{X_{i-j}} = \sqrt{X_i X_j}$;

—with the arithmetic mean coordinates $\mathbf{X_{i-j}} = \dfrac{X_i X_j}{2}$.

This recommendation should be used pragmatically, i.e., only in the case when it leads to the deeper minimum of the external accuracy criterion.

In another frequently occurring case, the cluster to be processed includes too many points, which drastically prolongs the processing time. Then we recommend evaluating all sample elements according to their efficiency and calculating the average efficiency of each row and column of the sample. Setting some threshold value for efficiency, we can exclude inefficient rows and columns from the sample, thus reducing it to the size determined by computer capabilities [10]. At the same time, by decreasing the threshold value, we increase not only the accuracy of the model, but also the calculation time. Vice versa, increasing the threshold value reduces both the accuracy and calculation time. Thus, we have a typical problem of the choice of optimal decision based on two criteria.

4.1. An Extended Combinatorial GMDH Algorithm with Simultaneous Optimization of the Numbers of a Sample's Rows and Columns Used in Modeling

The third degree of complexity characterizes GMDH algorithms which combine the optimization operations of the number of a sample's rows and columns used. This can be achieved by using the following formula for calculating the efficiency of all elements in the sample:

$$e = \frac{|(W - \overline{W})(V - \overline{V})|}{\max((W - \overline{W})(V - \overline{V}))},$$

where W is the initial value, $\overline{W}$ is its mean value, V is the value of the argument in the given sample element, and $\overline{V}$ is the mean value of the variable, to which the given sample element belongs.

Beginning with the value close to unity and gradually decreasing the threshold, we study the dependence of the external accuracy criterion from the value of the threshold. The minimum error corresponds to the optimal threshold value, which determines the optimal number of the used rows and columns of a data sample. Only a learning sample is subject to optimization. The test sample is used for calculating the external criterion. Generation of the models–candidates and their estimation is performed in the following order. If the efficiency threshold of sample elements is unity, the sample contains no elements, i.e., it is empty. We gradually decrease the threshold until at least one, the most efficient, variable appears in a sample which has several rows. Then, by using the first-row combinatorial GMDH algorithm (with the equation $y = a_0 + a_1x_i$), we find the simplest and most accurate model and determine the value of the external criterion by using the second-row combinatorial GMDH algorithm (with the equation $y = a_0 + a_1x_i + a_2x_j$). Next, by repeatedly reducing the threshold of efficiency, we find a model and a value of the third-row criterion and do this until the model's accuracy improves. The optimum model corresponds to the minimum of the external criterion. The extended GMDH algorithm is designed for use on large samples. For the small samples containing less than a hundred of rows, the models obtained by using extended and normal combinatorial algorithms usually coincide.

5. FOURTH STEP OF DESCENT: SELF-ORGANIZATION OF THE TWO- AND THREEFOLD MULTIROW NEURAL NETWORKS WITH ACTIVE NEURONS THAT WORK WITH GMDH ALGORITHMS

The fourth step of descent consists of the two- and threefold multirow neural networks with active neurons which are represented by the units working according the above-described GMDH algorithms. A neural network is a committee of the active neurons which differ in the output variable only. In a neural network, all the variables from the sample are regarded as output variables. If in this connection one variable, chosen in turn, is considered to be an output variable, it leads to the use of explicit templates. If all the variables are considered output, it leads to the use of implicit templates [10]. In the latter case, at each row of the net's self-organization, a system of linear algebraic equations should be solved in order to find the values of all output variables in this row.

The main problem of the fourth step of descent consists in the development of the theory and the algorithm of a neural net's self-organization, which is rather simple. By applying the combinatorial GMDH algorithm to each of the output variables, the first row of neurons is obtained (see Figs. 2 and 3). The first estimates of all the variables are obtained at the outputs of neurons of the first row; then, they are fed as additional secondary features or arguments at the second row of neurons, which also work according the combinatorial GMDH algorithm. The neuron's rows grow as long as the accuracy of the estimate of the variable under consideration increases.

5.1. Self-organization of the Architecture of the Twofold Multirow Neural Network

At the fourth step of descent, the twofold multirow neural network with the active neurons is self-organized; it was thoroughly described in [2]. The models, obtained at the previous steps of descent, can serve as the first row of the neural network. The output values of the first row are efficient arguments for the models of the second row, and the output values of the second row are included in the set of variables of the third row (Fig. 2). In this figure, the following notation is used: I-1, I-2, I-3, I-4,…, III-3, and III-4 are active GMDH neurons.

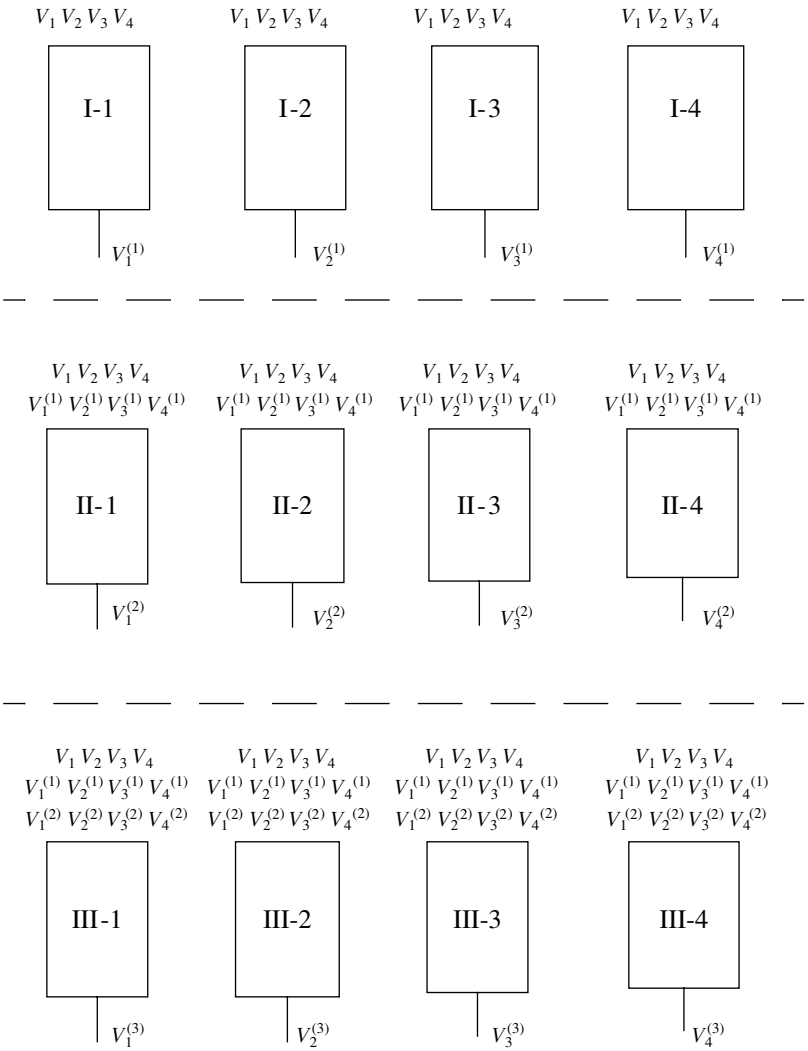


Fig. 2. The twofold multirow neural network with the active neurons constructed by GMDH algorithm for four variables.

The descent continues (i.e., the neural net's rows are growing) as long as the external criterion of accuracy decreases. As an alternative to the neural network constructed over the neurons which implement GMDH algorithms, we may suggest the construction of the neural network, which uses model-free analog algorithms [3].

5.2. Self-organization of the Threefold Multirow Neural Network

At the fourth step of descent, the problem of the stepwise prognosis of the processes requires two or more samples obtained on several similar objects. The prognosis model uses the variable from the sample, where this variable is predicted more accurately according to the value of external criterion. Here, for example, it turns out that for forecasting the efficiency of the New York Stock Exchange, the dollar rate should be forecasted on the basis of data from the New York Stock Exchange; sterling rate, from the London Stock Exchange, etc. This simple idea is implemented with the help of the threefold multirow neural network with active neurons, which are twofold multirow neural networks (Fig. 3), denoted as I-1, I-2, II-1, II-2, III-1, and III-2.

At this step of descent, the further extension of the search for efficient secondary arguments takes place. The descent continues until the external criterion decreases.

6. PARAMETRIC FILTERING (THE KALMAN FILTER TYPE) OF ADDITIVE NOISE RESISTANT TO FORMALIZING AND MEASURING BY GMDH ALGORITHMS

There are two consistent explanations of the efficiency of the neural network with active neurons. According to the first, a neural networks is treated as a generator of the efficient secondary features. According

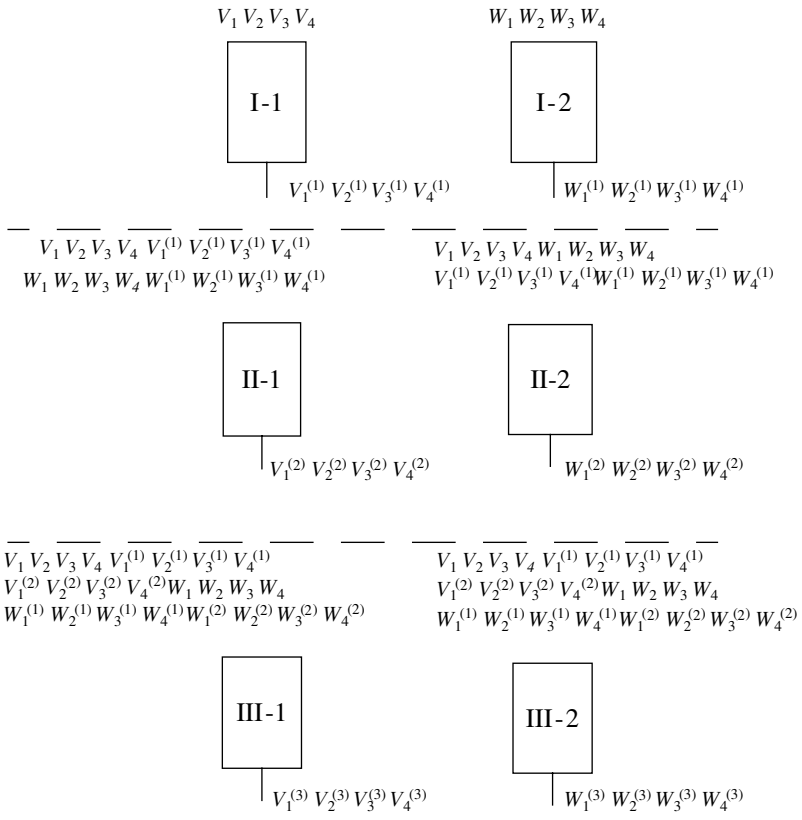


Fig. 3. The threefold multirow neural network with the active neurons being twofold multirow neural GMDH networks for four variables.

to the second, as a noise filter. The GMDH algorithms can be employed as parametric noise filters for obtaining the values of the variables indicated in the data sample.

The output variable indicated in the data sample can be considered as a sum of some true signal, set by the arguments indicated in the sample and algebraic sum of actions which were not included in the data sample [5]. This sum is called additive noise and can include ill-formalizable actions, e.g., psychological actions. If the noise variance is more than true signal variance, the dependence of external criterion upon the complexity of model structure is no longer minimal and GMDH algorithms do not work. To make them work, one should preliminary filter the additive noise.

Self-organization of the twofold multirow neural network is aimed at every possible reduction of every type of noise, including noise resistant to formalization and measurement.

It is known that noise filtration is actual in communication theory (frequency and synchronous filters) and in the automatic control systems where the estimates of the noisy output variables are obtained by the parametric Kalman filters [13, 14].

A combinatorial GMDH algorithm can be used as a parametric filter. In the twofold multirow neural network, each variable is processed along the chain of neurons, acting in accordance with GMDH algorithms. Moreover, in the chain, it is sufficient to search for the structure models by the external criterion in the last algorithm only, because only there a signal is considerably filtered from noise. The degree of filtering is controlled according to an external criterion. The number of rows for processing each element from the sample grows as long as the external accuracy criterion decreases. As a result, there can be an unequal number of rows for different variables in a neural network. Noisier variables demand more filtering rows [2].

Figure 4 exemplifies noise filtering by the use of a committee of twofold multirow neural networks in problems of multialternative pattern recognition. Here are some explanations to the figure:

1—The partitioning of data sample into k clusters. Data are normalized by the largest value of each variable. After the analysis of sample elements, inefficient rows and columns are excluded. A sample is partitioned into classes and clusters automatically or by experts by comparing two balance clusterization trees [1].

2—Various formations of samples for two-alternative recognition of each class (pattern).

3—The construction of a committee of twofold multirow neural networks, where the noise filtering is

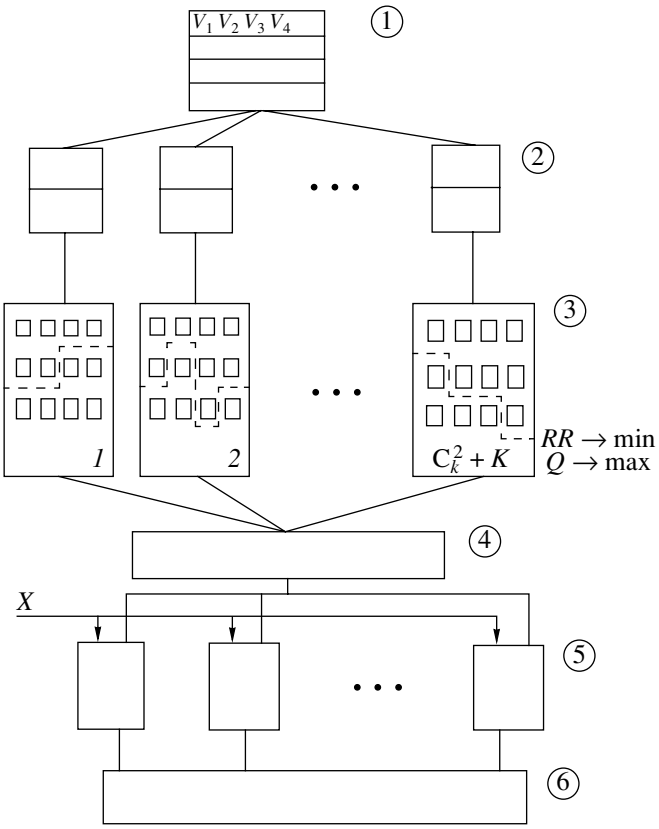


Fig. 4. The algorithm of multialternative pattern recognition with noise filtering:

1—the partitioning of data sample (performed automatically or by experts) into k clusters; 2—various formations of samples for two-alternative recognition; 3—the committee of twofold multirow neural networks; 4—the choice of the most accurate estimates and models; 5—the calculation of discriminant functions (models); 6—the comparison of the discriminant functions and decision making; X—input signal.

performed. At the outputs of neural networks, we obtain the discriminant functions for each class. The output value which indicates a membership of the feature vector in the given class is +1 for all vectors of the given class and –1 for all vectors of other classes. The number of rows in the lower and upper parts of the sample should be the same and contain the most efficient sample rows.

4—The choice of the most accurate estimates and models.

5—The calculation of discriminant functions.

6—The comparison of the discriminant functions.

X—The input signal which is received by all discriminant functions and belongs to the class whose function is closest to +1.

A committee of twofold multirow neural networks can be treated as a first row in a threefold multirow neural network outlined above.

7. MULTIALTERNATIVE PATTERN RECOGNITION

Figure 4 presents an algorithm which can serve as an example of multialternative pattern recognition. An input signal should be either referred to one of several patterns or rejected. Here, the various samples of two-alternative recognition are constructed along with the samples where the classes close according to some external criterion are combined into one general class. In the problems of multialternative pattern recognition, one should choose between the following two ways of designing the algorithm for a given problem.

(i) One can design all possible algorithms of two-alternative recognition and use the so-called voting [15, 16], where the recognition accuracy is achieved either if the number of patterns is small or if the voting has several rounds. In this case, an expert should vote at least two or three times.

(ii) One can rank all the features according to their efficiency and compose the new samples of two-alternative recognition which contain a set of points from each class and sets with the same number of points of all other classes (Fig. 4). A combinatorial GMDH algorithm yields a discriminant model for each sample. The input signal belongs to the class whose discriminant function is closest to +1.

Both algorithms of multialternative recognition were tested. However, we need to determine the field of application for each algorithm more thoroughly. If the voting is confined to one round, the determination coefficient, being an index of accuracy, rapidly decreases when the number of alternatives grows. In the second algorithm, the coefficient decreases slower. The determination coefficients are equal when the number of alternatives approximates seven.

8. AN EXAMPLE: THE ACTIVITY OF THE NEW YORK STOCK EXCHANGE FORECASTED BY USING DATA OF THE FIRST CLUSTER OF OPTIMAL PHYSICAL CLUSTERING AND ITS INTERMEDIATE GEOMETRICAL MEAN POINTS

By this example, we show that the accuracy of prognosis of the random process can be considerably increased by decreasing the number of rows in input data up to the vector set, defined by the first cluster of physical clustering, and by extending a sample with mean points. As an object, information from the New York Stock Exchange over the period May 11–July 17, 1991 was used.

The initial sample of experimental data contains four activity indices, shown in Table 1, where X_1 is the minimum value of activity index; X_2 is the maximum value of activity index; X_3 is the activity index at the close of the trading; and X_4 is a sum of total operations per day.

Table 1. Initial data sample

	X_1	X_2	X_3	X_4		X_1	X_2	X_3	X_4
1	375.0000	375.5000	376.5000	130.0000	36	371.0000	378.0000	378.0000	170.0000
2	371.0000	377.0000	372.0000	220.0000	37	377.0000	378.0000	378.0000	160.0000
3	366.0000	373.0000	368.0000	190.0000	38	372.0000	377.0000	373.0000	140.0000
4	369.0000	373.0000	373.0000	155.0000	39	374.0000	376.0000	375.0000	159.0000
5	369.0000	374.0000	373.0000	175.0000	40	372.0000	375.0000	374.0000	170.0000
6	372.0000	374.0000	373.0000	110.0000	41	371.0000	378.0000	378.0000	140.0000
7	373.0000	377.0000	376.0000	185.0000	42	376.0000	378.0000	376.5000	155.0000
8	375.0000	377.0000	377.0000	160.0000	43	376.0000	380.0000	376.0000	180.0000
9	374.0000	378.0000	375.0000	177.0000	44	376.0000	378.0000	377.5000	160.0000
10	375.0000	378.0000	377.0000	128.0000	45	376.0000	381.0000	380.0000	177.0000
11	378.0000	380.0000	379.0000	130.0000	46	380.0000	383.0000	383.0000	162.0000
12	377.0000	383.0000	383.0000	165.0000	47	381.0000	383.0000	382.0000	183.0000
13	382.0000	384.0000	383.0000	187.0000	48	381.0000	383.0000	381.0000	195.0000
14	383.0000	388.0000	387.0000	235.0000	49	381.0000	386.0000	386.0000	200.0000
15	385.0000	390.0000	390.0000	230.0000	50	384.0000	386.0000	385.0000	190.0000
16	387.0000	390.0000	388.0000	175.0000	51	382.0000	385.0000	383.0000	150.0000
17	385.0000	388.0000	388.0000	180.0000	52	379.0000	385.0000	379.5000	160.0000
18	384.0000	388.0000	385.0000	185.0000	53	378.0000	380.5000	380.3000	160.0000
19	383.0000	386.0000	384.0000	170.0000	54	380.0000	381.0000	381.0000	145.0000
20	379.0000	384.0000	380.0000	170.0000	55	380.0000	382.0000	381.0000	130.0000
21	387.0000	380.0000	379.0000	130.0000	56	380.0000	384.0000	384.0000	136.0000
22	379.0000	382.0000	381.0000	162.0000	57	384.0000	387.0000	387.0000	170.0000
23	374.0000	381.0000	376.0000	170.0000	58	387.0000	388.0000	388.0000	168.0000
24	375.0000	377.0000	377.0000	145.0000	59	387.0000	388.0000	387.6000	170.0000
25	377.0000	383.0000	383.0000	174.0000	60	387.0000	389.0000	387.5000	162.0000
26	380.0000	382.0000	380.0000	135.0000	61	385.0000	387.0000	385.5000	127.0000
27	378.0000	382.0000	378.0000	159.0000	62	384.0000	391.0000	391.0000	175.0000
28	375.0000	378.0000	376.0000	160.0000	63	390.0000	391.5000	390.0000	170.0000
29	374.0000	376.5000	375.0000	165.0000	64	388.0000	392.0000	389.0000	165.0000
30	375.0000	378.0000	378.0000	195.0000	65	387.0000	390.0000	387.5000	145.0000
31	371.0000	378.0000	371.0000	140.0000	66	386.0000	388.0000	388.0000	146.0000
32	360.0000	373.0000	371.0000	159.0000	67	388.0000	392.0000	390.0000	215.0000
33	368.0000	373.0000	372.0000	185.0000	68	389.0000	392.0000	390.0000	195.0000
34	372.0000	374.0000	374.0000	159.0000	69	389.0000	391.0000	389.0000	176.0000
35	368.0000	374.0000	371.0000	165.0000	70	385.0000	390.0000	395.0000	190.0000

Table 2 shows the sample in a single-moment form. For the forecast, we chose the 54th day of the work of the stock exchange. The forecasted value was chosen to be variable X_3.

To find optimal physical clustering, we used a computer program for calculating an *M*-type curve which expressed the dependence of the balance criterion from the number of construction step of the two classifica-

Table 2. Sample in the single-moment form

	Y_1	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}
3	373.00	366.00	373.00	368.00	190.00	371.00	377.00	372.00	220.00	375.00	375.50	376.50	130.00
4	373.00	369.00	373.00	373.00	155.00	366.00	373.00	368.00	190.00	371.00	377.00	372.00	220.00
5	373.00	369.00	374.00	373.00	175.00	369.00	373.00	373.00	155.00	366.00	373.00	368.00	190.00
6	376.00	372.00	374.00	373.00	110.00	369.00	374.00	373.00	175.00	369.00	373.00	373.00	155.00
7	377.00	373.00	377.00	376.00	185.00	372.00	374.00	373.00	110.00	369.00	374.00	373.00	175.00
8	375.00	375.00	377.00	377.00	160.00	373.00	377.00	376.00	185.00	372.00	374.00	373.00	110.00
9	377.00	374.00	378.00	375.00	177.00	375.00	377.00	377.00	160.00	373.00	377.00	376.00	185.00
10	379.00	375.00	378.00	377.00	128.00	374.00	378.00	375.00	177.00	375.00	377.00	377.00	160.00
11	383.00	378.00	380.00	379.00	130.00	375.00	378.00	377.00	128.00	374.00	378.00	375.00	177.00
12	383.00	377.00	383.00	383.00	165.00	378.00	380.00	379.00	130.00	375.00	378.00	377.00	128.00
13	387.00	382.00	384.00	383.00	187.00	377.00	383.00	383.00	165.00	378.00	380.00	379.00	130.00
14	390.00	383.00	388.00	387.00	235.00	382.00	384.00	383.00	187.00	377.00	383.00	383.00	165.00
15	388.00	385.00	390.00	390.00	230.00	383.00	388.00	387.00	235.00	382.00	384.00	383.00	187.00
16	388.00	387.00	390.00	388.00	175.00	385.00	390.00	390.00	230.00	383.00	388.00	387.00	235.00
17	385.00	385.00	388.00	388.00	180.00	387.00	390.00	388.00	175.00	385.00	390.00	390.00	230.00
18	384.00	384.00	388.00	385.00	185.00	385.00	388.00	388.00	180.00	387.00	390.00	388.00	175.00
19	380.00	383.00	386.00	384.00	170.00	384.00	388.00	385.00	185.00	385.00	388.00	388.00	180.00
20	379.00	379.00	384.00	380.00	170.00	383.00	386.00	384.00	170.00	384.00	388.00	385.00	185.00
21	381.00	387.00	380.00	379.00	130.00	379.00	384.00	380.00	170.00	383.00	386.00	384.00	170.00
22	376.00	379.00	382.00	381.00	162.00	387.00	380.00	379.00	130.00	379.00	384.00	380.00	170.00
23	377.00	374.00	381.00	376.00	170.00	379.00	382.00	381.00	162.00	387.00	380.00	379.00	130.00
24	383.00	375.00	377.00	377.00	145.00	374.00	381.00	376.00	170.00	379.00	382.00	381.00	162.00
25	380.00	377.00	383.00	383.00	174.00	375.00	377.00	377.00	145.00	374.00	381.00	376.00	170.00
26	378.00	380.00	382.00	380.00	135.00	377.00	383.00	383.00	174.00	375.00	377.00	377.00	145.00
27	376.00	378.00	382.00	378.00	159.00	380.00	382.00	380.00	135.00	377.00	383.00	383.00	174.00
28	375.00	375.00	378.00	376.00	160.00	378.00	382.00	378.00	159.00	380.00	382.00	380.00	135.00

Table 2. (Contd.)

	Y_1	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}
29	378.00	374.00	376.50	375.00	165.00	375.00	378.00	376.00	160.00	378.00	382.00	378.00	159.00
30	371.00	375.00	378.00	378.00	195.00	374.00	376.50	375.00	165.00	375.00	378.00	376.00	160.00
31	371.00	371.00	378.00	371.00	140.00	375.00	378.00	378.00	195.00	374.00	376.50	375.00	165.00
32	372.00	360.00	373.00	371.00	159.00	371.00	378.00	371.00	140.00	375.00	378.00	378.00	195.00
33	374.00	368.00	373.00	372.00	185.00	360.00	373.00	371.00	159.00	371.00	378.00	371.00	140.00
34	371.00	372.00	374.00	374.00	159.00	368.00	373.00	372.00	185.00	360.00	373.00	371.00	159.00
35	378.00	368.00	374.00	371.00	165.00	372.00	374.00	374.00	159.00	368.00	373.00	372.00	185.00
36	378.00	371.00	378.00	378.00	170.00	368.00	374.00	371.00	165.00	372.00	374.00	374.00	159.00
37	373.00	377.00	378.00	378.00	160.00	371.00	378.00	378.00	170.00	368.00	374.00	371.00	165.00
38	375.00	372.00	377.00	373.00	140.00	377.00	378.00	378.00	160.00	371.00	378.00	378.00	170.00
39	374.00	374.00	376.00	375.00	159.00	372.00	377.00	373.00	140.00	377.00	378.00	378.00	160.00
40	378.00	372.00	375.00	374.00	170.00	374.00	376.00	375.00	159.00	372.00	377.00	373.00	140.00
41	376.50	371.00	378.00	378.00	140.00	372.00	375.00	374.00	170.00	374.00	376.00	375.00	159.00
42	376.00	376.00	378.00	376.50	155.00	371.00	378.00	378.00	140.00	372.00	375.00	374.00	170.00
43	377.50	376.00	380.00	376.00	180.00	376.00	378.00	376.50	155.00	371.00	378.00	378.00	140.00
44	380.00	376.00	378.00	377.50	160.00	376.00	380.00	376.00	180.00	376.00	378.00	376.00	155.00
45	383.00	376.00	381.00	380.00	177.00	376.00	378.00	377.50	160.00	376.00	380.00	376.00	180.00
46	382.00	380.00	383.00	383.00	162.00	376.00	381.00	380.00	177.00	376.00	378.00	377.00	160.00
47	381.00	381.00	383.00	382.00	183.00	380.00	383.00	383.00	162.00	376.00	381.00	380.00	177.00
48	386.00	381.00	383.00	381.00	195.00	381.00	383.00	382.00	183.00	380.00	383.00	383.00	162.00
49	385.00	381.00	386.00	386.00	200.00	381.00	383.00	381.00	195.00	381.00	383.00	382.00	183.00
50	383.00	384.00	386.00	385.00	190.00	381.00	386.00	386.00	200.00	381.00	383.00	381.00	195.00
51	379.50	382.00	385.00	383.00	150.00	384.00	386.00	385.00	190.00	381.00	386.00	386.00	200.00
52	380.30	379.00	385.00	379.50	160.00	382.00	385.00	383.00	150.00	384.00	386.00	385.00	190.00
53	381.00	378.00	380.50	380.30	160.00	379.00	385.00	379.50	160.00	382.00	385.00	383.00	150.00
54	381.00	380.00	381.00	381.00	145.00	378.00	380.50	380.30	160.00	379.00	385.00	379.50	160.00

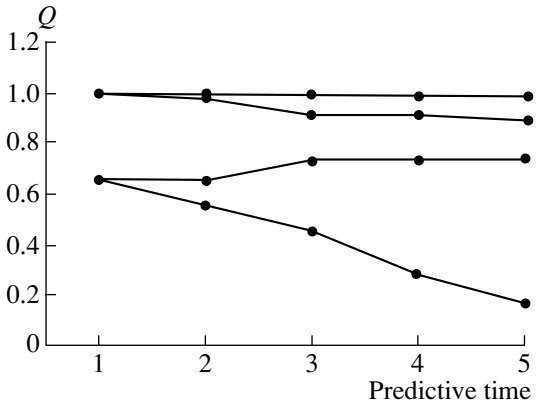


Fig. 5. Changing in determination coefficient.

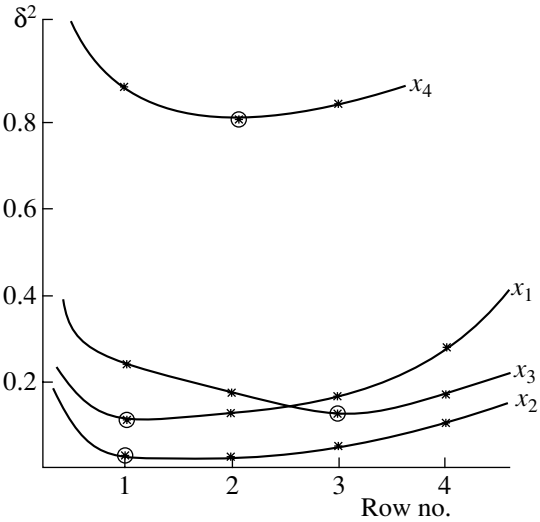


Fig. 6. Self-organizing of the twofold multirow neural network with the filtering of additive noise.

tion trees developed by A.P. Sarychev. The first cluster was proved to include the last twelve rows of Table 2.

By using the combinatorial GMDH algorithm, we add the mean geometric points for the first cluster and find the following forecasting models:

$$X_{3k+1} = 489.203 - 1.3271X_{1k} - 2.2648X_{2k} + 1.69759X_{3k} + 0.0771X_{4k} - 0.0258X_{1k-1} + 2.5930X_{2k-1} + 2.7702X_{3k-1} + 0.0012X_{4k-1} + 1.8739X_{1k-2} - 1.5382X_{2k-2} + 1.1435X_{3k-2} - 0.0472X_{4k-2}.$$

$$X_{3k+2} = 304.3073 - 2.0841X_{1k} + 0.4558X_{3k} + 0.0280X_{4k} + 1.9017X_{1k-1} - 1.3530X_{2k-1} + 1.9908X_{3k-1} + 0.1019X_{4k-1} + 1.0410X_{1k-2} - 0.5349X_{2k-2} - 1.1568X_{3k-2} - 0.2251X_{4k-2}.$$

$$X_{3k+3} = 328.0698 - 1.9519X_{1k} - 0.3918X_{2k} - 0.6967X_{3k} + 0.0385X_{4k} + 1.4974X_{1k-1} - 1.5971X_{2k-1} + 2.1631X_{3k-1} + 0.0811X_{4k-1} + 1.1755X_{1k-2} - 0.6985X_{2k-2} - 0.7023X_{3k-2} - 0.1941X_{4k-2}.$$

$$X_{3k+4} = 295.7382 - 1.6121X_{1k} + 0.3631X_{3k} + 0.0304X_{4k} + 1.8292X_{1k-1} - 1.2550X_{2k-1} + 1.5520X_{3k-1} + 0.0824X_{4k-1} + 1.0588X_{1k-2} - 0.5752X_{2k-2} - 1.0820X_{3k-2} - 0.1959X_{4k-2}.$$

$$X_{3k+5} = 366.3279 - 2.1159X_{1k} - 0.4593X_{2k} + 0.7830X_{3k} + 0.0408X_{4k} + 1.4909X_{1k-1} - 1.6150X_{2k-1} + 2.3175X_{3k-1} + 0.0814X_{4k-1} + 1.1249X_{1k-2} - 0.6836X_{2k-2} - 0.6733X_{3k-2} - 0.1975X_{4k-2}.$$

Figure 5 shows the change in the determination coefficient, which characterizes the accuracy of the forecast, while the predictive time gradually increases. Table 3 contains the prognostic values for the variables and the determination coefficient of the stepwise forecast in two versions:

(a) when the day of the forecast is constant;

Table 3. Prognostic and real values of X_3 and coefficient of determination

Prognostic days		55	56	57	58	59
(a)	For the entire table (54 points)	382.54	384.57	384.85	385.09	383.69
		0.6559	0.5561	0.4519	0.2796	0.1687
	For the first cluster with adding geometric means of the points	387.04	386.92	388.99	387.17	384.29
		0.9987	0.9799	0.9166	0.9125	0.8950
(b)	For the entire table (54 points)	382.54	385.55	387.55	386.46	386.86
		0.6559	0.6543	0.7259	0.7333	0.7380
	For the first cluster with adding geometric means of the points	387.04	387.78	387.26	386.46	386.86
		0.9987	0.9964	0.9961	0.9893	0.9819
Real values		387	388	387.6	387.5	385.5

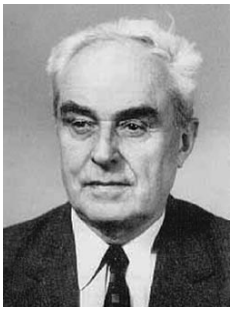
(b) when the time of forecast is shifted a step with each step of a forecast.

The forecast is made for all of Table 2 and for the first cluster with addition of the points with the geometric means of the coordinates.

By comparing the forecasts made for the whole sample and for the first cluster with the addition of the points with the geometric means of the coordinates, we make sure that by reducing the sample to the size of the first cluster and by introducing additional point, we can actually enhance the accuracy of the forecast. The coefficient of determination of the short-term forecast (one step forward) is increased from the value $Q = 0.6559$ to $Q = 0.8950$. For the intermediate-term forecast (five steps forward), the coefficient of determination is increased from the value $Q = 0.1687$ to $Q = 0.8950$, i.e., many times. Further enhancement of the accuracy can be achieved by using the Kalman filter by self-organization of the twofold multirow neural network (fourth step of descent). Figure 6 shows an example of the self-organizing neural network consisting of four neurons (borrowed from [2]), where $\delta^2 = (1 - \hat{x}_i)^2$. It is clear from the picture that different variables require the use of chains with a different number of neurons. In practice, instead of the additional points with the mean geometric coordinates, one can successfully use the additional points with the mean arithmetic coordinates.

REFERENCES

1. Ivakhnenko, A.G., Ivakhnenko, G.A., Savchenko, E.A., and Wunsh, 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.
2. Ivakhnenko, A.G., Ivakhnenko, G.A., and Mueller, J.-A., Self-organization of Neural Networks with Active Neurons, *Pattern Recognit. Image Anal.*, 1994, vol. 4, no. 2, pp. 185–196.
3. Ivakhnenko, A.G., Kovalishin, V.V., *et al.*, Application of Self-Organizing Neutral Networks with Active Neurons for Prediction of Bioactivity of Chemical Compounds by the Analog Search Algorithm, *J. of Automation and Information Sci.*, 1999, vol. 31, no. 7.
4. Shannon, C.E., *Works in Information Theory and Cybernetics*, Moscow: Inostrannaya Literatura, 1963.
5. Ivakhnenko, A.G. and Stepashko, V.S., *Pomekhoustoichivost' modelirovaniya* (Noise-Immunity of Modeling), Kiev: Naukova Dumka, 1985.
6. Krug, G.K. and Krug, O.Yu., Matematicheskii Metod Klassifikatsii Drevnei Keramiki (Mathematical Method for Classification of Ancient Ceramics) in *Trudy Instituta Arkheologii Akademii Nauk SSSR* (Proc. of Inst. of Archeology, Academy of Sciences of USSR), Moscow: Nauka, 1965, pp. 318–325.
7. Ivakhnenko, A.G., Ivakhnenko, G.A., Tetko, I.V., and Sarychev, A.P., Application of Analog Coordinates As Retarded Arguments of Virtual Processes of the Formation of Rows of a Data Sample, *Pattern Recognit. Image Anal.*, 1999, vol. 9, no. 3, pp. 401–407.
8. Ivakhnenko, A.G., *Samoobuchayushchiesya sistemy raspoznavaniya i avtomaticheskogo upravleniya* (Self-learning Systems of Recognition and Automatic Control), Kiev: Tekhnika, 1969.
9. Madala, H.R. and Ivakhnenko, A.G., *Inductive Learning Algorithms for Complex Systems Modeling*, Boca Raton: CRC, 1994.
10. Ivakhnenko, A.G. and Ivakhnenko, G.A., Problems of Further Development of GMDH Algorithms. Part 1, *Kibern. Vychisl. Tekh.*, 1999, issue 122.
11. Belogurov, V.P., A Criterion of Model Suitability for Forecasting Quantitative Processes, *Soviet J. Autom. Inform. Sci.*, 1990, vol. 24, no. 3, pp. 23–28.
12. Beer, S., *Cybernetics and Management*, London: English Univ., 1959.
13. Kalman, R.E. and Bucy, R., New Result in Linear Filtering and Prediction Theory, *J. Basic Eng.*, 1961, vol. 83, pp. 95–108.
14. Sage, A.P. and Melsa, J.L., *System Identification*, New York: Academic, 1971.
15. Aizerman, M.A., Braverman, E.M., and Rozonoer, L.I., *Metod potentsial'nykh funktsii v teorii obucheniya machin* (The Method of Potential Functions in the Theory of Machine Training), Moscow: Nauka, 1970.
16. Glaz, A.B., *Parametricheskaya i structurnaya adaptatsiya reshayushchikh pravil v zadachakh raspoznavaniya* (Parametric and Structural Adaptation of Decision Rules in Recognition Problems), Riga: Zinatne, 1988.



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.

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

Evgeniya A. Savchenko. Born 1969. Graduated from the Kiev Polytechnic 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.

Donald C. Wunsch. Received his BSci from the University of New Mexico in 1984, MSci from the University of Washington in 1987, and PhD (Electrical Engineering) in 1991 from the same university. In 1984–1993, a Senior principal scientist in the Boeing Company. In 1994–1999, an Associate Professor in Texas Technical University. Now, Distinguished Professor in the University of Missouri-Rolla (Computer Engineering). Published more than 100 technical papers on adaptive critic design, neural networks, fuzzy systems, reliability, nonlinear adaptive control, complex systems, financial engineering, and optimization.