Dollar Rate Forecasting by Group Method of Data Handling (GMDH) Algorithms and Accuracy Increase by Twice-Multilayered Networks

Ivakhnenko, Alexey Grigorevich
Glushkov Institute of Cybernetics, National Academy of Sciences of the Ukraina
252207 Kiev, Academica Glushkova 20, UKRAINA
E-mail: gai@runbox.com

Müller, Johann-Adolf
Fachbereich Informatik/Mathematik, Hochschule für Technik und Wirtschaft
D 01069 Dresden, F.-List-Platz 1, GERMANY
E-mail: Muellerj@informatik.htw-dresden.de

1. Backwards to probabilistic concept

At early stages of mathematical modeling we all believed that for solution of interpolation problems (such as function approximation, object identification, random processes and events forecasting, data sampling clusterization, pattern recognition, etc.) should be used probabilistic concept: optimal value of output variable should correspond to maximum of a posteriori probability calculated by Bayes formula or to the rules of Walds statistical decision theory [1]. Soon it was established that we have not in our disposition initial data samplings sufficient for reliable calculation of empirical probabilities necessary for these calculations. Even in the case when process is stationary there are necessary to have more than 30 lines (points or realizations) in the input data sampling [2]. We have often only 6 or 10 lines. That was reason why some scientists started to develop algorithms for number of lines artificial increase (bootstrap algorithm) and some authors developed polynomial GMDH algorithms, which can work on rather short data samplings beginning from 6-10 lines [3]. The problem remains to find algorithm which can work beginning from one line for each pattern (class) or discrete level of output variable [3].

Now we can say, that it was historically strategic mistake to refuse of probabilistic algorithms usage. We all simply have forgotten that except of usually used ordinary probabilities exist pairs probabilities which demands a very small number of data sampling lines for their calculation. Decision about optimal value of output variable should correspond to maximum of sum of ordinary and pairs empirical probabilities. To the number of ordinary random events E(y,xi) should be added the number of pairs random events E(y,xi,xj). Number of possible ordinary random events is equal to number of variables in data sample M. Number of possible pair random events is equal to 2^M.

2. *Example:* Rate of Dollar probabilistic forecasting

Let us consider algorithms for rate of Dollar stepwise forecasting. There are measured values of Dollar rate Y, rate of german Mark X_1, japanese Yen X_2, french Frank X_3 and Dow-Jones index X_4, for 60 days of observation. All the variables should be normalized by formulae:

$$x_i = (x_i - X_{i\,min})/(X_{i\,max} - X_{i\,min}) .$$

Units of measurement for all the variables are changed to receive equal coefficients (equal to one) in regression model:

$$y = a_0 + x_1 + x_2 + x_3 + x_4 .$$

The interval of output variable y (normalized rate of Dollar) change is divided to 11 discrete levels $y = R_1$, $y = R_2$,..., $y = R_{11}$. For each level its individual data sampling is gathered, from all input data, containing equal number of points.

For example, for level R_3 we have following data sampling: days of level R_3 (I,J,K,L) observation and normalized values of variables measured one day before forecast are shown in the Table 1.

Table 1. Input data for level R_3 (empirical probabilities or number of random events calculation).

	x_1	x_2	x_3	x_4
I	5	4	3	2
J	3	5	4	3
K	5	5	3	2
L	5	5	3	2

Using this local data sampling we calculate ordinary and pairs empirical probabilities to reach level R_3 next to the day when there were: $x_1=5$, $x_2=4$, $x_3=3$ and $x_4=2$.(example only of forecast output vector). Sum of ordinary empirical random events, proportional to ordinary empirical probabilities, are shown in table 2.There are shown four variables x_2 x_3 x_4 and x_5 and their levels 2, 3, 4 and 5 which take part in given forecast output vector.

Table 2. Sum of ordinary random events.

	2	3	4	5
x_1	0	1	0	3
x_2	0	0	1	3
x_3	0	3	1	0
x_4	3	1	0	0

Sum of ordinary random events $\sum P_{ord} = 16$.

Number 1 shows that events are calculated with delay in 1 day, as usually for stepwise forecast. Sum of pairs empirical random events are shown in table 3. There are shown all possible pairs of variables: their discrete values (called discreets) and sums of pairs random events.

Table 3. Sums of pair random events to reach possible discreets.

	5-4	5-3	5-2	4-3	4-2	3-2
x_1-x_2	3	3	0	1	0	0
x_1-x_3	3	3	0	1	0	0
z_1-x_4	0	0	3	0	0	3
x_2-x_3	3	3	0	3	0	0
x_2-x_4	0	3	3	1	3	0
x_3-x_4	0	0	0	1	3	3

Sum of pairs of random events $\sum P_{pair} = 49$.

Criterion:

$$\sum P = \sum P_{ord} + \sum P_{pair} \rightarrow \max,$$

should be calculated for all levels of output variable R_1 R_2....R_{11} and the most probable levels should be chosen out as the most probable forecast for each given output vector of factors.

2.1. Taking prehistory of process into account

Only to show brief example we did not take into account delayed arguments with lag longer than one day. To get high accuracy of forecast there is necessary to take into account random events calculated for each delay time separately. For example for Dollar rate forecast it is necessary to take into account 35 delayed values of Dollar rate. Criterion in following form would be calculated:

$$\sum P = \sum_{1}^{35} \sum \left(P_{i\,ord} + P_{i\,pair} \right) \rightarrow \max$$

2.2. Comparison with polynomial GMDH algorithm

In the problem considered we have in our disposition rather long input data sampling and there-fore we can apply polynomial GMDH [3] too. The result is almost the same as by probabilistic algorithm: average error of forecast with lead time of eight days is equal to one percent, i.e. is very small one.[4]. Lead time of forecast can be estimated approximately by formulae:

$$T = \frac{h}{RR}; \qquad RR = \frac{\sum_{i=1}^{T_L}(y_i - y_{iF})^2}{\sum_{i=1}^{T_L}(y_i - \bar{y})^2},$$

where: h - time step given in data sample, RR-error variance criterion
(when $RR = 0.01$ lead time is $T_L = 100$ time steps).

3. Transformation of initial data sampling into sampling of centers of clusters of physical clusterization as the tool for long data samplings

There are two different difficulties in models self-organization by GMDH algorithms. First arises in case when we receive too short initial data sampling. The use of pair probabilities is recommended in this case. They can be applied when we have in data sampling only one line for each division (level) of output variable. Second difficulty arises when we receive too long initial data sampling. In this case can be recommended to provide clusterization of the data sampling, to find so called physical clusterization and to use for modelling the sampling of clusters center coordinates [5].

Let us give necessary definitions: each line in input data sampling corresponds to coordinates of same point in multi-dimensional characteristic space. Clusters are the compact groups of points in this space. Clusterization means the division of this space to clusters. Each cluster can be presented by its points and by average point called as center of cluster [5]. Physical or optimal clusterization can be defined as such for which the number of clusters and coordinates of their centers are stable: they can be found on each long enough part of initial data sampling. This property allows us to recommend physical clusterization as criterion to choose optimal forecasting among many forecastings-candidates. Physical clusterization can be considered as discrete analogue of continue-valued physical model of object.

In case of a very long initial data sampling we can recommend to use for modelling the short sampling of clusters centers coordinates. This sampling is usually very short and only modelling algorithms with pairs probability account can be applied. Number of lines is equal to number of clusters. Type of algorithm which should be applied depends from dispersion of noise in the input data sampling and from its size: from number of lines N and from number of variables (factors) M. When noise dispersion is small (less 20-30%), polynomial parametric algorithms are most effective. For very small dispersion usual internal criteria can be used. It is not necessary to divide data sampling to two parts.

When we have from 6 to 20 lines in the sampling Combinatorial GMDH algorithm is recommended. When number of factors is equal from 21 to about 500 lines the multilayered GMDH algorithms can be used [3]. But when number of lines is big the algorithm of data sampling clusterization should be used, to transform initial data sampling into sampling of centers of clusters of physical clusterization coordinates.

Genetic types of algorithms, with crossover recombination of "mother" and "father" partial descriptions and random mutation of new models-candidates are recommended in cases when data sampling contains more than 500 variables. If we say that it is enough to have one line of data for each level of output variable it means from genetic approach we use algorithm of fern plants, having only "mother."

4. Neural network with active neurons as the tool to increase accuracy of interpolation problems solution [4]

Each single algorithm of modelling can be considered as one big active neuron which find by learning or self-organization process optimal input factors and therefore solves the problem of network self-organization. To construct first layer of neurons of neural network it is enough to use active neurons for all the variables given in input data sampling. It was proved experimentally and theoretically that output variables of active neurons of first layer are a very effective secondary factors for the active neurons of second and next layers, especially in case of noised input data. Number of layers should be increased until the internal error criterion decreases. Procedure of divergence-convergence of factors number, along the network layers, allows to choose the most effective regression space [6].

5. Neural network with probabilistic active neurons purpose of which is to find "true” input clustering

Neural networks for solution of interpolation problems use for their self-organization error criterion. Problem of physical model of the object reconstruction and correction of mistakes in input data samplings can be solved by another type of networks working on convergence criterion. Output of active neurons of first layer is transformed into special second data sampling and all calculations are repeated. Algorithm is iterational one. Number of active neurons, working consequently, increased, until the data sampling converges to one stable "true" form [7]. If initial data sampling is given by an expert then "true", clustering corresponds to majority of expert decision in each particular case solved by him.

6. Future development of neural networks is construction of committee of decision rules committees

Committee of decision rules is the union of several modeling algorithms each having its “area of competence“. Algorithms differ mainly by their reference functions [8]. Change of situation is reflected in data sampling and for each situation should be used one the most effective algorithm. Such committee can be considered as one big active neuron of neural network. This network can be considered as committee of committees.

References

1. Wald A. Sequential Analysis. London, Charman and Hall Ltd. - 1940.- p.327.
2. Draper N.Smith H. Applied regression analysis. New York, - 1966. - p.310.
3. Madala H. R. and Ivakhnenko A.G. Inductive Learning Algorithms for Complex Systems Modelling. CRC press Inc, - 1994. - p.388.
4. Ivakhnenko A.G. and Müller J.A. Self-organization of nets with active neurons, SAMS, 20, 1995, no.1-2, pp.93-106.
5. Ivakhnenko A.G. and Müller J.A Self-organization of optimal physical clustering of data sampling for weakened description and forecasting of fuzzy objects. Pattern Recognition and Image Analysis, 3, 1993, no.4, pp. 417-422.
6. Ivakhnenko G.A. Self-organization of neuronet with active neurons for effects of nuclear tests explosion forecasting. SAMS, 20, 1995, no. 1-2, pp. 107-117.
7. Ivakhnenko A.G., Petukhova S.A. et all. Objective Selection of Optimal Clustering of a Data Sampling During Compensation of Non-Robust Interference. J. of Automation and Information Science, 26, 1993, no.3, pp.45-56.
8. Mazurov V.D., Krivonogov A.I. at all. Committees in Decision Making. Kibernetika, 1984, no.1, pp. 90-95.
9. Articles at http://GMDH.come.to