Hybrid GMDH Deep Learning Networks - State-of Art and New Prospective Trends

Yuriy Zaychenko[a] and Galib Hamidov[b]

[a] *Institute for Applied System Analysis, Igor Sikorsky Kyiv Polytechnic Institute, Peremogy avenue 37, Kyiv, 03056, Ukraine*
[b] *Information Technologies Department, Azershig,str. K. Kazim-zade 20, Baku, , AZ1008, Azerbaijan*

Abstract

In this paper new class of deep learning (DL) neural networks is considered and investigated-so-called hybrid DL networks based on self-organization method Group Method of Data Handling (GDMDH). The application of GMDH enables not only to train neural weights but to construct the network structure as well. As nodes of this structure different elementary neurons with two inputs may be used. So, the advantage of such structure is a small number of tuning parameters. In the paper the following types of neurons are considered: Wang-Mendel network with two inputs and neo-fuzzy neurons. The advantage of the neo-fuzzy neurons is unlike general fuzzy neurons is absence of fuzzy membership functions training and less computational time for training. The application of GMDH enables to train neuron weights sequentially layer after layer in the process of construction network structure until the stop criterion holds. Such approach allows to exclude drawbacks of DL training algorithms -decay or explosion of gradient. The process of structure construction and optimization using GMDH algorithm is presented. The numerous applications of suggested hybrid DL networks for solution of AI problems like forecasting of share prices and market indicators at various stock exchanges are considered and analyzed. The comparison with conventional DL networks is performed which enables to estimate their efficiency and advantages.

Keywords [1]

Hybrid deep learning networks, self-organization, structure optimization, forecasting

1. Introduction

Last years deep learning (DL) networks are widely used in different problems of artificial intelligence: forecasting, pattern recognition, medical diagnostics, etc.[1-4]. For its training various algorithms were developed usually based on Back propagation method. Presence of many layers when using gradient algorithm usually lead to occurrence drawbacks as vanishing or explosion of gradient. Therefore, the approach was suggested how to exclude this drawback to perform layer after layer training using stacked encoder-decoder or stacked restricted Boltsmann machines [1, 2]. However, the problem is left how to choose the number of layers in DL network. The existing methods of DL don't enable to generate structure of Dl networks. But the training process will be more efficient if to adapt not only neuron weights but the structure of network as well. For this goal the application of GMDH method seems very promising. GMDH is based on principle of self- organization and enables to construct network structure automatically in the process of algorithm run [5-7]. In the previous years GMDH-neural networks having active neurons [5-7], R-neurons [19], Q-neurons [3] as nodes were developed; in the area integrating fuzzy GMDH and neural networks the GMDH-neuro-fuzzy systems and GMDH-neo-fuzzy systems [13] were developed; GMDH-wavelet-neuro-fuzzy systems [14,15] were also elaborated. The very important property of GMDH is that as building blocks for construction of a structure of DL networks elementary models with only two inputs, so-called

Information Technology and Implementation (IT&I-2021), December 01–03, 2021, Kyiv, Ukraine
EMAIL: zaychenkoyuri@ukr.net (A. 1); galib.hamidov@gmail.com (A. 2)
ORCID: 0000-0001-9662-3269 (A. 1); **0000-0002-9942-1950** (A. 2)

partial descriptions, are used. This allows to cut substantially training time for hybrid Dl network as compared with conventional DL networks.

Therefore, GMDH-hybrid neuro-fuzzy system was developed in [16] that combines advantages of the traditional GMDH and DL fuzzy networks and may be trained with simple learning procedures. The nodes of this network are Wang-Mendel elementary neural networks with only two inputs. The experimental investigations of this class of hybrid DL networks have shown their efficiency and preference over conventional DL networks. But the drawbacks of application of Wang-Mendel networks as nodes of hybrid Dl networks lies herein that it's necessary to train not only neural weights but membership functions as well.

Therefore, later another class of hybrid networks – GMDH – neo-fuzzy networks were developed wherein as nodes of network -neo-fuzzy neurons with two inputs are used [17]. For their training its necessary to adapt only neuron weights that demands less computational resources and cuts training time. That's very important for DL networks with a large number of hidden layers. The experimental investigations of hybrid -neo-fuzzy networks and comparison with conventional DL network have shown their efficiency and less computational calculations for training.

The goal of this paper is to investigate different hybrid GMDH-neo-fuzzy networks with small number of adjusted parameters and estimate their efficiency for structure optimization and forecasting.

2. Hybrid network structure optimization based on GMDH method

The GMDH method was used to synthesize the structure of the hybrid network based on the principle of self-organization. The successive increase in the number of layers is carried out until the value of the external criterion of optimality MSE begins to increase for the best model of the current layer. In this case it is necessary to return to the previous layer, to find there the best model with the minimum value of criterion. Then we move backward, go through its connections, find the corresponding neurons of the previous layer. This process continues until we reach the first layer and the corresponding structure is automatically determined.

The process of synthesis of the network structure in the forward direction is shown in Fig. 1 where in green color the outputs which passed through selection block (SB)are shown while in red color - outputs which were dropped (excluded) by SB.

The process of restoring the desired structure in the backward direction is shown in Fig. 2. In the yellow color nodes and their connections selected by this process are indicated

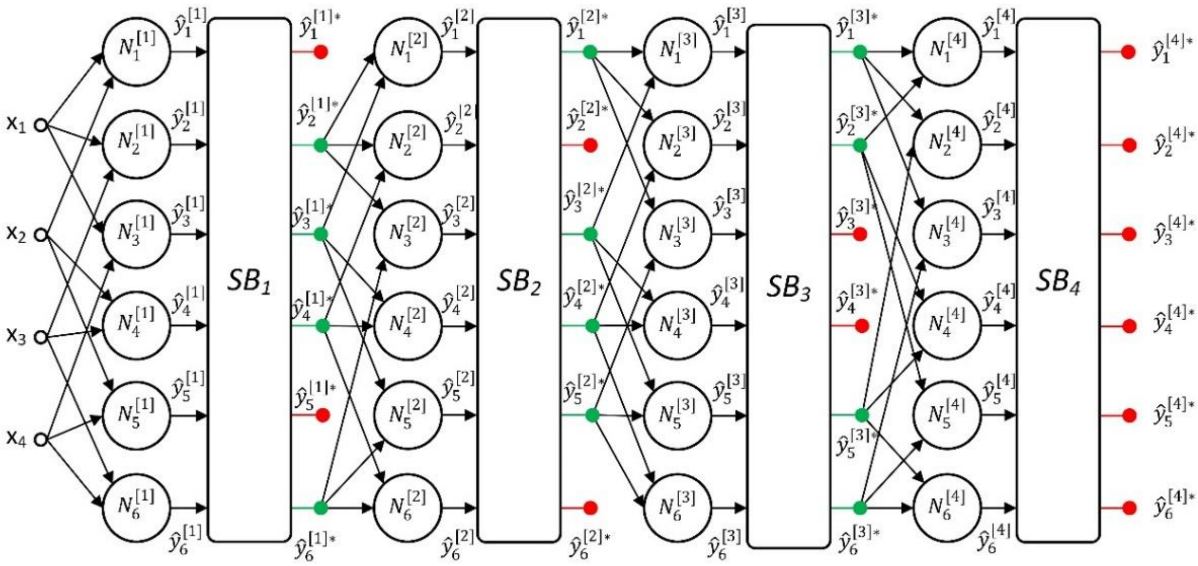

Figure 1. Hybrid network structure construction using GMDH method

The corresponding optimal constructed structure of the hybrid network for this forecasting problem is shown in Fig. 3. It consists of 3 layers: first layer has 3 neo-fuzzy neurons, second layer- two neurons and the last- one neuron.

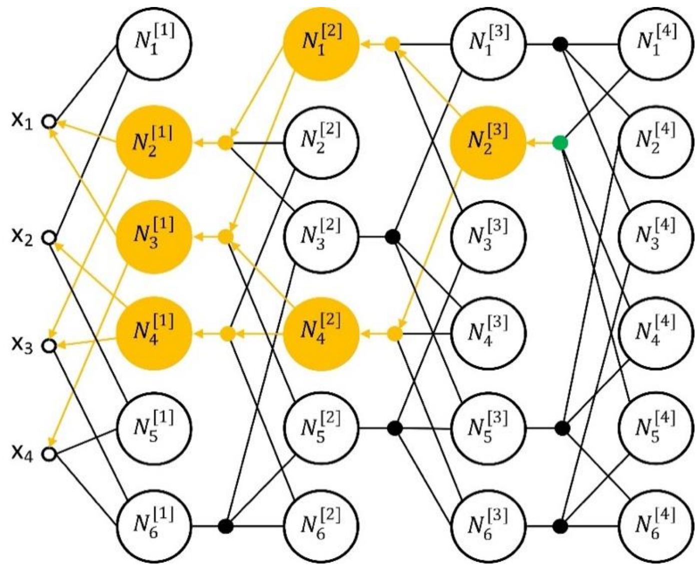

Figure 2. Process of restoring found optimal structure in backward direction

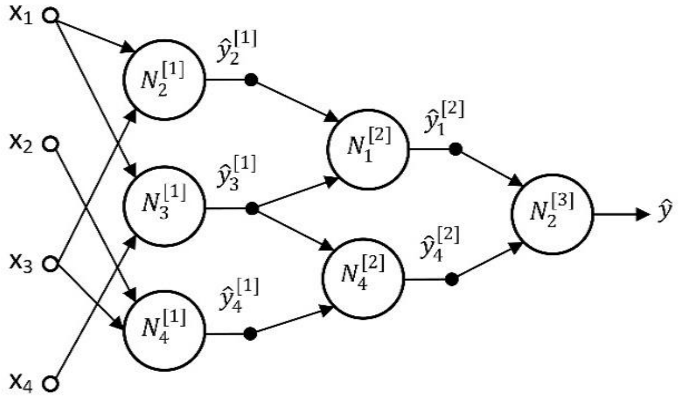

Figure 3. Optimal Structure of hybrid network for covid forecast constructed by GMDH

3. Experimental investigations of hybrid GMDH -fuzzy networks in forecasting problems

For efficiency estimation of hybrid GMDH DL networks the problems of the forecasting share price and market indices at the stock exchanges were considered. The experimental investigations for stock prices forecasting were carried out. In the first experiment as a forecasted variable the RTS index in 2013 with time step one week was chosen. As external regressors (inputs) stock prices of the leading companies were used. Total sample had 55 points that was used while searching the optimal partial description in the GMDH.

As the accuracy criteria of the obtained models MAPE and RMSE were used. In the first experiment the dependence of MAPE on inputs number was explored. The forecasting results for hybrid neuro-fuzzy network are presented in the table 1. For comparison the corresponding results for full cascade neo-fuzzy network (NFN) network are presented.

As it follows, hybrid GMDH-neuro-fuzzy network has higher accuracy than the cascade neuro-fuzzy network due to properties of hybrid networks.

In the next experiment the problem of forecasting share prices of Microsoft corp was considered. As input sample the stock prices of Microsoft corp. since 01.11.14 to 29.12.14 were used. The sample size was 64 points. The training sample included 62 points the test sample 4 points. The forecasting interval was for 4 steps ahead, the first two steps are checked with available data. The constructed GMDH-neuro-fuzzy network had 6 fuzzy inputs. The experimental results are presented in Table 2 and Table 3.

As it follows from the experimental results the GMDH-neuro-fuzzy network showed better forecasting accuracy than the cascade neuro-fuzzy network. Its MAPE value doesn't exceed 0.4%.

Table 1.

Accuracy for hybrid GMDH-network and Cascade neo-fuzzy network

number of inputs	MAPE for hybrid GMDH-network	MAPE for cascade NFN
2	0,04038	0,06031
4	0,03950	0,05141
6	0,03998	0,04425
8	0,04248	0,04396
10	0,04935	0,05171
12	0,04084	0,04465

Table 2.

Forecasting Results for hybrid GMDH- network

Date	Real value	Predicted value	Absolute error	Relative error, %
26.12.14	18030,2	17971,63	58,577	0,325
24.12.14	18053,7	17991,94	61,772	0,342

Table 3.

Forecasting Results (MAPE) for Different Neuro-Fuzzy Networks and GMDH

Real value	GMDH-neuro-fuzzy network	GMDH system	Cascade-neuro-fuzzy network
48.14	0.623	1. 20	3.40
47.88	2.13	1.94	2.54
average	1.377	1.57	2.97

In the next experiment training time for GMDH-neuro-fuzzy network, and cascade fuzzy network were compared. In the Table 4 the training time in seconds for hybrid GMDH- neuro-fuzzy network and full cascade neuro-fuzzy network is presented. As an initial sample Microsoft stock prices in the same period since 01.11.14 to 29.12.14 was used.

Table 4.

Training time for hybrid GMDH network and cascade network

Inputs number	GMDH-hybrid network, s	cascade network, s
2	0,004	0,015
4	0,009	0,021
6	0,013	0,037
8	0,021	0,048
10	0,030	0,053

In the next experiments efficiency of hybrid neo-fuzzy network in forecasting index NASDAQ was explored. The data was taken in the period from 13.11. 17 till 29.11.19. The sample size was 510 points. As an output variable the closing price of the index NASDAQ next day was taken. In the first experiment the accuracy dependence on number of inputs for hybrid neo-fuzzy network was investigated. In the table 5 the forecasting results are presented under different inputs number with 8 membership functions per variable (parameter h) and ratio training/test =70/30. In the next experiment the investigation of error dependence on number of MF per variable (parameter h) was performed. Number of inputs was n=5, training/test ratio was 70/30. The results are presented in Fig.4.

Analyzing these results, one may conclude that with MF number rise MAPE first falls, then attains minimum and after then begins to rise. That fully matches to self- organization principle of GMDH method [3]. The best value was obtained with the following parameters values: number of inputs n=5, h=8, number of layers 4 and MAPE value is 3,91.

Table 5.
Forecasting MAPE versus inputs number for hybrid neo-fuzzy netwrok

Inputs number	2	3	4	5	6	7	8	9	10
MAPE	5,2	4,7	4,33	3,91	4,22	4,72	5,24	5,53	5,85

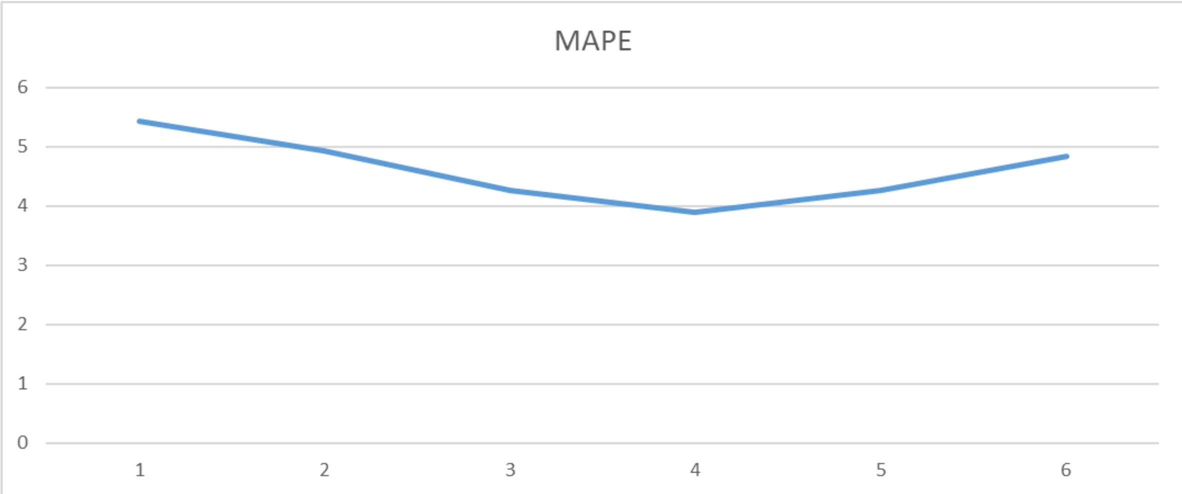

Figure 4. MAPE versus number of membership functions h per variable

For forecasting efficiency estimation of the hybrid network, it was compared with a cascade neo-fuzzy network [11] and GMDH at the same data. In the cascade neo-fuzzy network, the following parameters values were used: number of inputs n=9, number of rules 9, cascades number is 3. The comparative forecasting results are presented in the table 6, training sample – 70%.

Analyzing these results one can easily conclude the suggested hybrid neo-fuzzy network and neuro-fuzzy network have the best accuracy, the second one is GMDH method and the worst is the cascade neo-fuzzy network. The forecasting accuracy of both hybrid networks differs insignificantly.

In the next experiments the training time of different hybrid networks and alternative NN was investigated and compared. In the table 7 the training time in seconds for GMDH-neuro-fuzzy and -neo-fuzzy network and full cascade neuro-fuzzy network are presented. As an initial sample we used Microsoft stock prices in the period since 01.11.14 to 29.12.14., a sample size is 64 points.

As it follows from the presented results the least training time has hybrid neo-fuzzy network, the second place takes hybrid neuro fuzzy network and the last is full cascade network

4. Optimization of hybrid GMDH -neo-fuzzy network in the problem of forecasting

In the next experiments investigations of hybrid GMDH -neo-fuzzy network in the problem of Dow Jones Index forecasting were performed and compared with FNN ANFIS.

The Dow Jones is the stock index of the 30 largest American companies, which was founded in 1896. The initial data was taken from Yahoo, a leading financial information provider owned by Yahoo! To prepare the initial data, data were uploaded at various intervals, namely the value of the stock index by days, weeks and months. Each of the sets contains the following data:

- Date - data period;
- Open - opening price;
- High - the highest price for the period;
- Low - the lowest price for the period;
- Close - the price at the end of the period;

- Adj Close - average closing price;
- Volume - sales for the period.

Table 6.
MAPE values for different forecasting methods

inputs number/method	Hybrid neuro-fuzzy network	Hybrid GMDH- neo-fuzzy network	GMDH	Cascade neo-fuzzy neural network
4 inputs	4,30	4,31	4,19	6,04
5 inputs	3,93	3,91	4,11	6,09
6 inputs	4,35	4,36	5,53	8,01
7 inputs	4.80	4,77	6,26	8,68

Table 7.
Training time for different fuzzy neural models

Inputs number	Time for GMDH-neuro-fuzzy network, s	Time for GMDH-neo-fuzzy network, s	Time for full cascade network, s
2	0.004	0.003	0.015
4	0.009	0.007	0.021
6	0.013	0.012	0.037
8	0.021	0.018	0.048
10	0.030	0.025	0.053

The data set for the interval of one day contains 4867 records, of which non-zero records are 4788 ones. The data set for the interval one month contains 1001 records, of which 1000 records are non-zero. The data set for the interval of one month contains 195 records, of which 195 are non-zero.

Data normalizing. Reduction to a single scale is provided by normalization of each variable to the range of its values. In the simplest case, it is a linear transformation

$$\mathrm{m}\overline{x_i} = \frac{x - x_{i\,min}}{x_{i\,max} - x_{i\,min}} \text{ in the interval } x_i \in [0, 1].$$

To find the most informative features as an input vector, we have alternately trained the network on data sets that transmit only the following features subsets:

('Open', 'High', 'Low', 'Volume', 'Close'); ('Open', 'High', 'Low', 'Volume');
('Open', 'High', 'Low', 'Close'); ('Open', 'High', 'Low'); ('Open', 'High', 'Close');
('Open', 'High', 'Volume'); ('Open', 'Close', 'Low'); ('Open', 'Volume', 'Low');
('High', 'Low', 'Close'); ('Open', 'High'); ('High', 'Close'); ('Low', 'Close'); ('Open', 'Volume').

The main network parameters that can be configured include the size of the input vector, the number of rules, and the function that sets them, the number of parameters that are transferred to the next layer.

The size of the input vector is determined by the number of informative features that are transmitted for training, and the number of days on the basis of which the network gives the predicted value. Also, the number of network functions that can be set includes the number of membership functions and their appearance, as well as the degree of freedom of choice of the system.

To select these parameters, it is necessary to conduct an experiment, training the system, setting these parameters in the interval, and keeping those that give the best results in the test sample.

The following parameters were investigated:

- n – number of preceding days, based on which the forecasting is performed (sliding window size). N ∈ [1; 6];
- h - number of membership functions in each node, h ∈ [2; 9];
- s – membership function parameter $exp\left[-\left(\frac{x-c_i}{2\sigma}\right)^2\right]$, where $\sigma = \frac{(b-a)}{h}(s * (h - 1))$;

- b – an interval end;
- a – an interval beginning;
- h – membership functions number, which cover the interval;
- $s \in [0.01;\ 1.5]$;
- $f -$ number of parameters which are transferred to the network next layer (freedom of choice).

To set of initial data was divided into a training sample and test sample in the ratio of 70% and 30%, respectively. Having launched GMDH-neo-fuzzy system for training, values of MAE and MAPE criteria were obtained with different combinations of these parameters. For the Dow Jones stock index with different forecast intervals, the best parameters for the different set of informative features were obtained as a result of training and testing, which are shown in Table 8.

Table 8.
The results of the selection of the optimal parameters of GMDH-neo-fuzzy system for Dow Jones index with different prediction intervals

Sets of informative features	1 month						1 week					
	n	h	f	s	MAE	MAPE	n	h	f	s	MAE	MAPE
'Open', 'High', 'Low', 'Volume', 'Close'	1	2	2	1.0	**0.0147**	**0.0452**	2	4	2	0.7	**0.0077**	**0.0295**
'Open', 'High', 'Low', 'Volume'	1	2	3	1.3	0.0156	0.0476	2	4	3	0.9	0.0086	0.0332
'Open', 'High', 'Low', 'Close'	1	2	2	1.0	**0.0147**	**0.0453**	2	4	2	0.7	**0.0077**	**0.0295**
'Open', 'High', 'Low'	1	2	3	1.3	0.0156	0.0476	2	4	3	0.9	0.0086	0.0332
'Open', 'High', 'Close'	1	2	3	1.2	0.0153	0.0467	2	4	3	0.9	0.0079	0.0309
'Open', 'High', 'Volume'	5	2	5	0.1	0.0177	0.0654	2	4	3	1.0	0.0098	0.0380
'Open', 'Low', 'Close'	1	2	3	1.2	0.0147	0.0456	2	4	3	0.7	0.0081	0.0308
'Open', 'Volume', 'Low'	5	3	7	0.1	0.0171	0.0644	4	2	6	0.1	0.0095	0.0348
'High', 'Low', 'Close'	1	2	2	1.0	**0.0147**	**0.0453**	2	4	2	0.7	**0.0077**	**0.0295**
'Open', 'High'	5	2	5	0.1	0.0177	0.0654	2	4	3	1.0	0.0098	0.0380
'Open', 'Close'	1	2	2	1.3	0.0165	0.0498	2	4	3	0.6	0.0085	0.0331
'High', 'Close'	1	2	2	1.2	**0.0154**	**0.0467**	2	4	3	0.9	0.0079	0.0309
'Low', 'Close'	1	2	2	1.2	**0.0147**	**0.0456**	2	4	2	0.7	0.0081	0.0306
'Open', 'Volume'	5	2	2	0.8	0.0189	0.0689	3	4	2	0.1	0.0112	0.0445

Thus, analyzing presented results one may conclude that the most informative for GMDH-neo-fuzzy system are the following sets of features: ['Open', 'High', 'Close'], ['Open', 'Low', 'Close'], ['High', 'Low', 'Close'], ['High', 'Close'], ['Low', 'Close'].

For the Dow Jones stock index for one month forecast period, the following optimal configurations of GMDH-neo-fuzzy network were obtained:

- the number of informative features - 3;
- the number of periods on the basis of which the forecast is made - 1;
- the number of membership functions in each of the nodes - 2;
- the number of layers - 2;
- the number of nodes in the first layer – 3;
- number of nodes on the second layer – 1.

For the Dow Jones stock index for the one week forecast period, the following optimal configurations of the GMDH-neo-fuzzy system were obtained:

- the number of informative features- 3;
- the number of periods on the basis of which the forecast is made - 2;
- the number of membership functions in each of the nodes - 4;

- the number of layers - 2;
- the number of nodes on the first layer - 15;
- the number of nodes on the second layer – 1.

The form of the membership function for forecasting interval of one week is shown in the Figure 5.

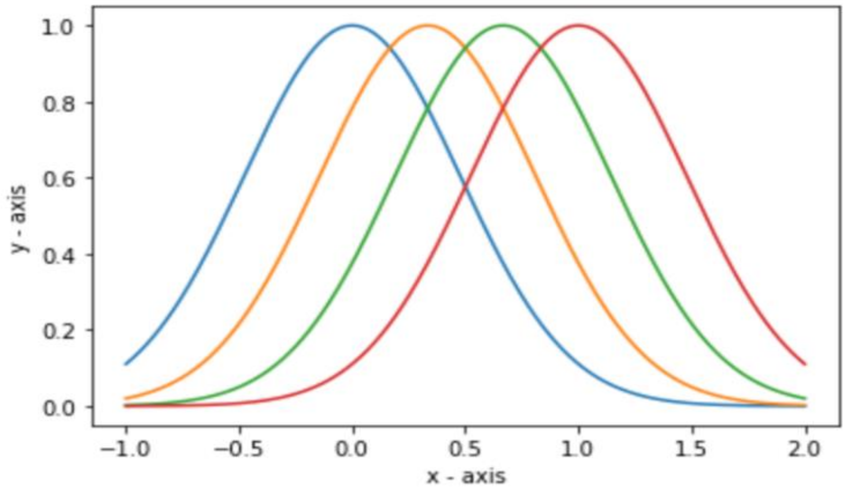

Figure 5. Forms of the membership function of Dow Jones index for the forecast period of 1 week

For the Dow Jones stock index for one week forecast period, the following optimal configurations of GMDH-neo-fuzzy network were obtained:

- number of informative features - 3;
- the number of periods on the basis of which the forecast is made - 5;
- the number of membership functions in each of the nodes - 2;
- the number of layers - 2;
- the number of nodes in the first layer - 105;
- the number of nodes in the second layer - 1

Next, experiments were performed to find the optimal values of the parameters of FNN ANFIS. The size of the input vector is determined by the number of informative features that are transmitted for training, and the number of days of prehistory, on the basis of which the forecasting is performed.

To select these parameters, an experiment was performed, including training of the network, setting these parameters in the interval, and choosing those that give the best results at the test sample.

The following intervals for parameters were set:

- n is the number of previous days on the basis of which the forecast is made (the size of the sliding window), $n \in [1; 6]$;
- h - the number of membership functions in each of the nodes, $h \in [2; 9]$.

The set of initial data was divided into a training sample and test data in the proportion of 70% and 30%, respectively. By launching the ANFIS network with different combinations of these parameters, data on MAE and MAPE criteria were obtained.

For the Dow Jones stock index one month forecast period, the following optimal ANFIS network configurations were obtained:

- number of informative features - 3;
- number of nodes – 6;
- the number of periods on the basis of which the forecast is made - 2;
- the number of membership functions in each of the nodes - 6.

After finding all the optimal parameters of GMDH-neo-fuzzy system and training parameters, the system was trained, and then the data for prediction was provided. Training and testing of the system took place on data for the period up to 01.01.2021 for monthly periods, and until 01.06.2021 for weekly and day periods. Forecasting was based on data for the period after 01.01.2021. for monthly periods and after 01.06.2021 for day and week periods. For Dow Jones index with a forecast period of one month, the following forecasting data were obtained: MAE - 0.02952; MAPE - 0.0335, forecasting time - 0.00025s. Learning and forecasting results are shown in Figure 6.

5. Comparison of forecasting results of GMDH-neo-fuzzy system and ANFIS network

Experimental investigations of the accuracy of Dow Jones index forecasting with forecasting intervals of one month, one week and one day were performed. using a hybrid GMDH-neo-fuzzy network. For each prediction interval the optimal parameters found in previous experiments were selected. A comparative analysis with the forecasting results obtained by FNN ANFIS was performed. According to the results of forecasting, values of MAE, MAPE and training time for each type of neural network were obtained. All comparison results are summarized in Tables 10 – 12.

Table 9.
The results of the selection of the optimal characteristics of ANFIS network for Dow Jones index with different forecast intervals

Sets of informative features	1 month				1 week				1 day			
	n	h	MAE	MAPE	n	h	MAE	MAPE	n	h	MAE	MAPE
'Open', 'High', 'Low'	2	6	0.222	0.0710	1	9	0.0091	0.0334	1	10	0.0037	0.0142
'Open', 'High', 'Close'	2	3	0.0223	0.0727	2	8	0.0080	0.0303	1	11	**0.0034**	**0.0129**
'Open', 'Low', 'Close'	2	6	**0.0192**	**0.0680**	2	10	0.0804	0.0307	1	5	0.0045	0.0154
'High', 'Low', 'Close'	2	8	0.0209	0.0720	2	9	0.0903	0.0325	2	10	0.0036	0.0134
'High', 'Close'	2	9	0.0223	0.0750	1	3	**0.0077**	**0.0282**	1	7	0.0035	0.0135
'Low', 'Close'	2	7	0.0201	0.0691	1	5	0.0094	0.0338	1	5	0.0035	0.0136

As one can see, for all forecasting intervals the best forecasting results were obtained for hybrid GMDH-neo-fuzzy system. The worst forecasting result for ANFIS network was obtained for one month forecasting period. The largest difference in the accuracy of forecasting by both criteria was obtained for the forecasting period of one month (over 200%). As the forecasting period decreases, the gap between the networks accuracy also decreases. In addition, training and direct prediction times were also significantly less for hybrid GMDH-neo-fuzzy system.

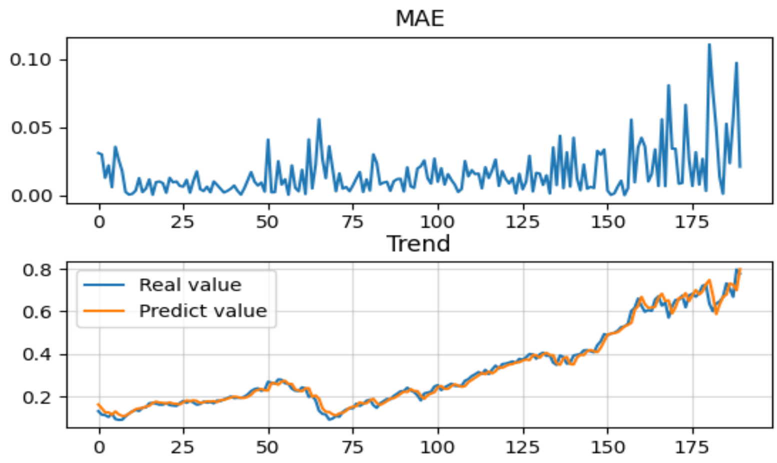

Figure 6. Results of training and forecasting Dow Jones Index with interval one month by hybrid GMDH neo-fuzzy system

Table 10.
Comparison of the forecasting results of GMDH-neo-fuzzy neural network and FNN ANFIS for Dow Jones Index with forecasting interval 1 month

Criterion	GMDH-neo-fuzzy neural network	Мережа ANFIS	Difference
MAE at training sample	0.016938	0.016135	4.70%
MAPE at training sample	0.061866	0.052607	14.97%
MAE at test sample	0.02952	0.096734	-227.68%
MAPE at test sample	0.03350	0.107397	-220.59%
Training time (sec.)	0.0023246	75.258	32375x
Forecasting time (sec)	0.0003123	0.02652	84.92x

Table 11.
Comparison of the forecasting results of GMDH-neo-fuzzy neural network and FNN ANFIS for Dow Jones Index with forecasting interval 1 week

Criterion	GMDH-neo-fuzzy neural network	FNN ANFIS	Difference
MAE at training sample	0.007949	0.008564	-7.74%
MAPE at training sample	0.029890	0.029291	2.00%
MAE at test sample	0.011476	0.019279	-67.99%
MAPE at test sample	0.012468	0.020923	-67.82%
Training time (sec.)	0.012840	194.3520	14980x
Forecasting time (sec)	0.00027132	0.028604	105.42x

Table 12.
Comparison of the forecasting results of GMDH-neo-fuzzy neural network and FNN ANFIS for Dow Jones Index with forecasting interval 1 day

Criterion	GMDH-neo-fuzzy neural network	FNN ANFIS	Difference
MAE at training sample	0.003618	0.004234	-17.03%
MAPE at training sample	0.013981	0.014067	-0.615%
MAE at test sample	0.005348	0.005822	-8.86%
MAPE at test sample	0.005812	0.005822	-0.172%
Training time (sec.)	0.19944	876.3658	4394.13x
Forecasting time (sec)	0.00040317	0.038055	94.39x

6. Conclusion

In the paper hybrid GMDH -neuro-fuzzy and neo-fuzzy networks are considered and investigated.

The algorithm of hybrid network structure synthesis is presented and demonstrated at the problem of forecasting. The experimental investigations of the hybrid networks were carried out and compared with conventional DL networks. The experiments have shown that forecasting accuracy of hybrid neuro-fuzzy and neo- fuzzy networks at the considered problems are approximately equal and it's better than for alternative DL cascade neo- fuzzy networks and GMDH. The problem of forecasting Dow Jones Index with application of hybrid neo- fuzzy networks was considered, investigated and compared with FNN ANFIS at the different forecasting intervals. The optimal parameters of hybrid neo- fuzzy networks were found. The experimental results have shown the forecasting accuracy of hybrid neo-fuzzy networks is much better than for FNN ANFIS. The training time is the least for hybrid neo- fuzzy networks as compared with all considered alternative DL networks.

In a whole the hybrid DL networks based on GMDH are free from drawbacks of conventional DL networks- decay or explosion of gradient. Besides, they enable to construct optimal network structure automatically in the process of algorithm GMDH run and additionally they demand less computational costs for training due to small number of tunable parameters (only two) in every hidden node as compared with DL networks of general structure. That's is especially significant for DL networks with large number of layers.

7. References

[1] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[2] G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural Computation, vol. 18, no. 7, pp. 1527–1554, May 2006.

[3] Y. Bengio, Y. LeCun, and G. Hinton, "Deep learning," Nature, no. 521, pp. 436–444, May 2015.

[4] J. Schmidhuber, "Deep learning in neural networks: an overview," Neural Networks, no. 61, pp. 85-117, 2015.

[5] A.G. Ivakhnenko, G.A. Ivakhnenko, J.A. Mueller, Self-organization of the neural networks with active neurons. Pattern Recognition and Image Analysis 4, 2 (1994): 177- 188.

[6] A.G. Ivakhnenko, D. Wuensch, G.A. Ivakhnenko, Inductive sorting-out GMDH algorithms with polynomial complexity for active neurons of neural networks. Neural Networks 2 (1999): 1169-1173.

[7] G.A. Ivakhnenko, Self-organization of neuronet with active neurons for effects of nuclear test explosions forecasting. System Analysis Modeling Simulation 20 (1995): 107-116.

[8] M. Zgurovsky, Yu. Zaychenko, Fundamentals of computational intelligence- System approach. Springer, 2016.

[9] L.-X. Wang and J. M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning". IEEE Trans. on Neural Networks 3, №5 (1992): 807-814.

[10] J.-S. Jang, "ANFIS: Adaptive-network-based fuzzy inference systems". IEEE Trans. on Systems, Man, and Cybernetics. 23 (1993,): 665-685.

[11] T. Yamakawa, E. Uchino, T. Miki, H. Kusanagi, A neo-fuzzy neuron and its applications to system identification and prediction of the system behavior, in: Proc. 2nd Intern. Conf. Fuzzy Logic and Neural Networks «LIZUKA-92». Lizuka, 1992, pp. 477-483.

[12] Ye. Bodyanskiy, N. Teslenko and P. Grimm, Hybrid evolving neural network using kernel activation functions, in: Proc. 17th Zittau East-West Fuzzy Colloquium, Zittau/Goerlitz, HS, 2010, pp. 39 46.

[13] Ye. Bodyanskiy, Yu. Zaychenko, E. Pavlikovskaya, M. Samarina and Ye. Viktorov, The neo-fuzzy neural network structure optimization using the GMDH for the solving forecasting and classification problems, Proc. Int. Workshop on Inductive Modeling, Krynica, Poland, 2009, pp. 77-89.

[14] Ye. Bodyanskiy, O. Vynokurova, A. Dolotov, O. Kharchenko, Wavelet-neuro-fuzzy network structure optimization using GMDH for the solving forecasting tasks, in: Proc. 4th Int. Conf. on Inductive Modelling ICIM 2013, Kyiv, 2013, pp. 61-67.

[15] Ye. Bodyanskiy, O. Vynokurova and N. Teslenko, Cascade GMDH-wavelet-neuro-fuzzy network, in: Proc. 4th Int. Workshop on Inductive Modeling «IWIM 2011», Kyiv, Ukraine, 2011, pp. 22-30.

[16] Ye. Bodyanskiy, O. Boiko Yu. Zaychenko, G. Hamidov, Evolving Hybrid GMDH-Neuro-Fuzzy Network and Its Applications, in: Proceedings of the Intern conference SAIC 2018, Kiev, Ukraine, 2018.

[17] Evgeniy Bodyanskiy, Yuriy Zaychenko, Olena Boiko, Galib Hamidov, Anna Zelikman. The hybrid GMDH-neo-fuzzy neural network in forecasting problems in financial sphere, in: Proceedings of the International conference IEEE SAIC 2020, Kiev, Ukraine, 2020.

[18] D. T. Pham and X. Liu, Neural Networks for Identification, Prediction and Control. London, Springer-Verlag, 1995.

[19] T. Ohtani, Automatic variable selection in RBF network and its application to neuro-fuzzy GMDH, Proc. Fourth Int. Conf. on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000, V.2, pp. 840-843