Forecasting Model Accuracy Enhancement Using Trend Extraction

Ievgeniia Savchenko, Gregory Ivakhnenko, Liudmyla Somina

International Research and Educational Center of Information Technologies and Systems of NAS of Ukraine, pr. Academika Glushkova, 40, Kyiv, 03680, Ukraine

savchenko@irtc.org.ua, gau@gmdh.net

Abstract. *The proposed method is based on calculation of a regular trend deviation. It can be applied to the forecast model accuracy enhancement in regular processes. The method consists in calculation of regular trend after real data sample and forecasting of the trend deviation by combinatorial GMDH algorithm. An example of blood glucose level forecasting in the task of diabetic treatment is shown.*

Keywords

Combinatorial GMDH algorithm, trend, regular process, forecasting, blood glucose level.

1. Introduction

The example of optimal model search for the forecasting of blood glucose level (BGL) of diabetic treatment after real clinical data [1] using combinatorial Group Method Data Handling (GMDH) algorithm [2] is shown in the paper.

Three ways of optimal model building for BGL forecasting are compared:

1) by nonlinear autoregression model;
2) by time series trend;
3) by sum of trend and trend deviation forecast.

Data of monitoring diabetic treatment are supplied by scientists of Hungarian Logic Laboratory [3]. BGL of patient and dose of insulin injection are measured four times per day: at 6, 12, 17 and 22 hours. It is required to find the forecast of BGL for three days forward. That gives the doctor a possibility to predict the illness evolution and correct the patient treatment.

2. Nonlinear autoregression model construction (without the trend selection)

At first the forecasting model algorithm is building by combinatorial GMDH. Such models can be optimal for the stationary time series. Input data sample of monitoring diabetic treatment are presented in the Table 1.

Variables in the table mean the following:

x_i - blood glucose level at 6, 12, 17, and 22 hours.

z_i - doses of insulin injection at 6, 12, 17, and 22 hours.

Two new variables are added to the data sample:

$$w_1 = x_i \cdot z_i \,;\; w_2 = \frac{x_i}{z_i} \,. \tag{1}$$

It is required to find the nonlinear autoregression model:

$$x_{i\,k+1} = f\,(x_{i\,k}, z_{i\,k}, w_{1k}, w_{2k}) \,,\; i = 6,12,17,22 \,. \tag{2}$$

Table 1. Input data sample of monitoring diabetic treatment

Data	x_{6k}	x_{12k}	x_{17k}	x_{22k}	z_{6k}	z_{12k}	z_{17k}	z_{22k}
16.08.1993	15.4000	12.4000	9.7000	7.9000	20.0	10.0	14.0	12.0
17.08.1993	10.5000	10.6000	5.5000	7.0000	20.0	12.0	14.0	12.0
18.08.1993	13.5000	24.6000	4.4000	13.6000	20.0	12.0	14.0	12.0
19.08.1993	14.9000	16.4000	5.8000	13.6500	20.0	10.0	14.0	12.0
20.08.1993	22.5000	11.4000	6.1000	13.7000	20.0	12.0	14.0	12.0
21.08.1993	10.0000	18.5000	18.1000	19.1000	20.0	12.0	12.0	12.0
22.08.1993	17.2000	14.0000	5.8000	6.4000	19.0	12.0	12.0	12.0
23.08.1993	16.3000	8.5000	7.9000	11.3000	19.0	11.0	14.0	12.0
24.08.1993	15.2000	11.5000	8.0000	6.8000	19.0	9.0	12.0	12.0
25.08.1993	14.0000	21.7000	24.4000	9.8000	19.0	11.0	14.0	12.0
26.08.1993	21.1000	10.4000	11.0000	15.3000	17.0	11.0	14.0	12.0
27.08.1993	14.2000	24.7000	20.7000	10.2000	19.0	11.0	14.0	12.0
28.08.1993	14.3000	19.6000	16.0000	9.6000	18.0	11.0	14.0	12.0
29.08.1993	26.9000	28.1000	20.9000	16.7000	18.0	11.0	13.0	12.0
30.08.1993	26.9000	24.5000	16.5000	7.2000	18.0	11.0	13.0	12.0
31.08.1993	16.5000	16.8000	19.6000	10.9000	18.0	10.0	13.0	12.0
01.09.1993	16.3000	21.4000	16.7000	13.0000	18.0	10.0	13.0	12.0
02.09.1993	15.4000	20.5000	20.4000	20.6000	17.0	10.0	13.0	12.0
03.09.1993	13.7000	15.8000	17.4000	14.4000	17.0	10.0	13.0	12.0
04.09.1993	17.0000	14.6000	17.7000	20.0000	15.0	10.0	13.0	12.0
05.09.1993	9.6000	12.1000	16.4000	12.1000	17.0	10.0	13.0	12.0

A fragment of input data sample for forecasting BGL by combinatorial GMDH algorithm is shown in the Table 2.

Table 2. Data sample for modeling by combinatorial GMDH

Data	x_{ik} ***, mmol/l***	i	w_1	w_2	x_{ik+1} ***,mmol/l***
17.08.1993	10.500	20.000	210.000	0.525	13.500
	10.600	12.000	127.200	0.883	24.600
	5.500	14.000	77.000	0.393	4.400
	7.000	12.000	84.000	0.583	13.600
18.08.1993	13.500	20.000	270.000	0.675	14.900
	24.600	12.000	295.200	2.050	16.400
	4.400	14.000	61.600	0.314	5.800
	13.600	12.000	163.200	1.133	13.650
19.08.1993	14.900	20.000	298.000	0.745	22.500
	16.400	10.000	164.000	1.640	11.400
	5.800	14.000	81.200	0.414	6.100
	13.650	12.000	163.800	1.138	13.700
20.08.1993	22.500	20.000	450.000	1.125	10.000
	11.400	12.000	136.800	0.950	18.500
	6.100	14.000	85.400	0.436	18.100
	13.700	12.000	164.400	1.142	19.100
21.08.1993	10.000	20.000	200.000	0.500	17.200
	18.500	12.000	222.000	1.542	14.000
	18.100	12.000	217.200	1.508	5.800
	19.100	12.000	229.200	1.592	6.400

Optimal nonlinear autoregression model:

$$\hat{x}_{i\,k+1} = 13{,}021 + 2{,}99\,w_{2k}\,. \qquad (3)$$

The values of criteria: $AR = 3{,}61$; $BS = 1{,}82$,

where AR is the regularity criterion; BS is the error bias criterion [4].

The real data of BGL and calculated are presented in the Fig. 1.

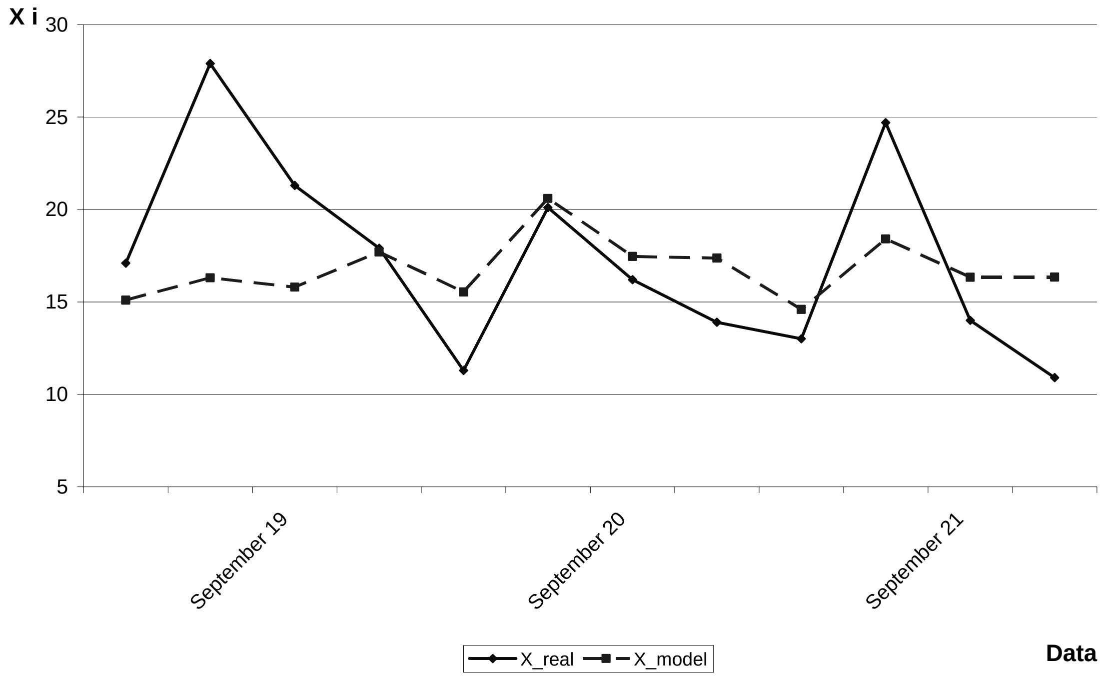

Figure 1. Real BGL data and calculated by combinatorial GMDH algorithm

Results of modeling by combinatorial GMDH algorithm are presented in the Table 3.

Table 3. Results of modeling by nonlinear autoregression model

№	Data of September	t	x_{ik}, ***mmol/l***	$\hat{x}_{ik}$, ***mmol/l***	Error of model, %
1	19	6	11.697	11.697	11.697
2		12	41.581	41.581	41.581
3		17	25.800	25.800	25.800
4		22	1.089	1.089	1.089
5	20	6	37.517	37.517	37.517
6		12	2.497	2.497	2.497
7		17	7.790	7.790	7.790
8		22	24.963	24.963	24.963
9	21	6	12.269	12.269	12.269
10		12	25.473	25.473	25.473
11		17	16.678	16.678	16.678
12			49.895	49.895	49.895
Mean error for three days of forecasting					***21.438***

3. Trend model construction

Let us introduce the following trend function:

$$\hat{x}_{i\,tr} = a_0 + a_1 T + a_2 t + a_3 Tt + a_4 T^2 + a_5 t^2 , \tag{4}$$

where T - day number, t - time (hours).

Fragment of data sample for value of trend function search is presented in the Table 4. Days was numbered from "–34" to "–1" for the data sample which was used for the model construction and from "1" to "3" for the days of forecasting.

Table 4. Fragment of data sample for forecast by trend

№	Data	T	t	$T \cdot t$	T^2	t^2
1	*17.08.1993*	-34.000	6.000	-204.000	1156.000	36.000
2		-34.000	12.000	-408.000	1156.000	144.000
3		-34.000	17.000	-578.000	1156.000	289.000
4		-34.000	22.000	-748.000	1156.000	484.000
5	*18.08.1993*	-33.000	6.000	-198.000	1089.000	36.000
6		-33.000	12.000	-396.000	1089.000	144.000
7		-33.000	17.000	-561.000	1089.000	289.000
8		-33.000	22.000	-726.000	1089.000	484.000
9	*19.08.1993*	-32.000	6.000	-192.000	1024.000	36.000
10		-32.000	12.000	-384.000	1024.000	144.000
11		-32.000	17.000	-544.000	1024.000	289.000
12		-32.000	22.000	-704.000	1024.000	484.000
…	…	…	…	…	…	…
125	*18.09.1993*	-3.000	6.000	-18.000	9.000	36.000
126		-3.000	12.000	-36.000	9.000	144.000
127		-3.000	17.000	-51.000	9.000	289.000
128		-3.000	22.000	-66.000	9.000	484.000
129	*19.09.1993*	-2.000	6.000	-12.000	4.000	36.000
130		-2.000	12.000	-24.000	4.000	144.000
131		-2.000	17.000	-34.000	4.000	289.000
132		-2.000	22.000	-44.000	4.000	484.000
133	*20.09.1993*	-1.000	6.000	-6.000	1.000	36.000
134		-1.000	12.000	-12.000	1.000	144.000
135		-1.000	17.000	-17.000	1.000	289.000
136		-1.000	22.000	-22.000	1.000	484.000

Equation of trend describes regular part of process of medical treatment of diabetes. The estimations of coefficients from find by data sample of Table 4 with using a least-squares method.

The following trend function was obtained:

$$\hat{x}_{i\,tr} = 8{,}69 - 0{,}23\,T + 1{,}45\,t + 0{,}017\,Tt - 0{,}0006T^2 - 0{,}0047t^2 . \tag{5}$$

The BGL forecast is calculated by equation (5) for three days from 19 to 21of September (Table 5).

Table 5. Result of modelling by trend

№	Data	T	t	x_{ik}, *mmol/l*	$\hat{x}_{ik}$, *mmol/l*	Model error, %
1	19	1.0	6	15.568	17.10	8..959
2		1.0	12	19.240	27.90	31.040
3		1.0	17	19.679	21.30	7.609
4		1.0	22	17.736	17.90	0.915
5	20	2.0	6	15.434	11.30	36.586
6		2.0	12	19.208	20.10	4.438
7		2.0	17	19.732	16.20	21.803
8		2.0	22	17.874	13.90	28.589
9	21	3.0	6	15.299	13.00	17.685
10		3.0	12	19.175	24.70	22.370
11		3.0	17	19.784	14.00	41.312
12		3.0	22	18.010	10.90	65.232
Mean error of three days of forecasting						***23.878***

4. Combined (trend + trend deviation) model

The value of trend deviation can be calculated with the goal to increase the forecasting model accuracy.

$$\Delta x_i = x_{i\,real} - x_{i\,tr}\,, \tag{6}$$

where $x_{i\,real}$ - real BGL data, $x_{i\,tr}$ - BGL value obtained by trend function.

Let us determine the value Δx_i (6) to find a trend value by the described above algorithm.

Firstly the trend deviation is forecasted by combinatorial GMDH algorithm. Then BGL forecast is finding as sum of trend value and forecast of trend deviation.

The following model is obtained by combinatorial GMDH algorithm:

$$\Delta\hat{x}_{i\,k+1} = 1{,}54 + 2{,}082 w_{2k} - 0{,}29 x_{i\,tr\,k}\,. \tag{7}$$

Values of criteria: $AR = 0{,}228$; $BS = 0{,}0178$.

Results of BGL forecasting for the three days are presented in the Table 6 including the combined model:

$$\hat{x}_{i\,k+1} = \hat{x}_{i\,tr\,k+1} + \Delta\hat{x}_{i\,k+1}\,. \tag{8}$$

Table 6. Results of BGL forecasting

№	Data of September	x_{ik}, *mmol/l*	$x_{i\,tr\,k+1}$, *mmol/l*	$\Delta x_{i\,k+1}$ *Mmol/l*	$\Delta\hat{x}_{i\,k+1}$, *mmol/*	$\hat{x}_{i\,k+1}$, *mmol/l*	Error of model, %
1	19	17.10	14.352	-0.34	-1.08	14.513	16.07
2		27.90	20.513	0.07	1.31	20.578	26.48
3		21.30	18.809	-0.29	-0.92	18.790	11.70
4		17.90	17.453	-0.32	-0.42	17.346	2.49
5	20	11.30	13.638	-0.19	-1.66	13.799	20.69
6		20.10	18.998	0.45	-0.18	19.064	5.48
7		16.20	18.119	-0.47	-1.66	18.102	11.85
8		13.90	11.926	-0.58	-6.08	11.820	14.20
9	21	13.00	16.504	0.05	1.34	16.666	26.95
10		24.70	18.470	-0.78	-0.67	18.538	25.22
11		14.00	17.955	-0.40	-1.88	17.939	28.25
12		10.90	12.956	0.09	-5.19	12.852	18.86
Mean error of three days of forecasting							***17,35***

Results of modeling by adding the trend deviation are presented in the Figure 2.

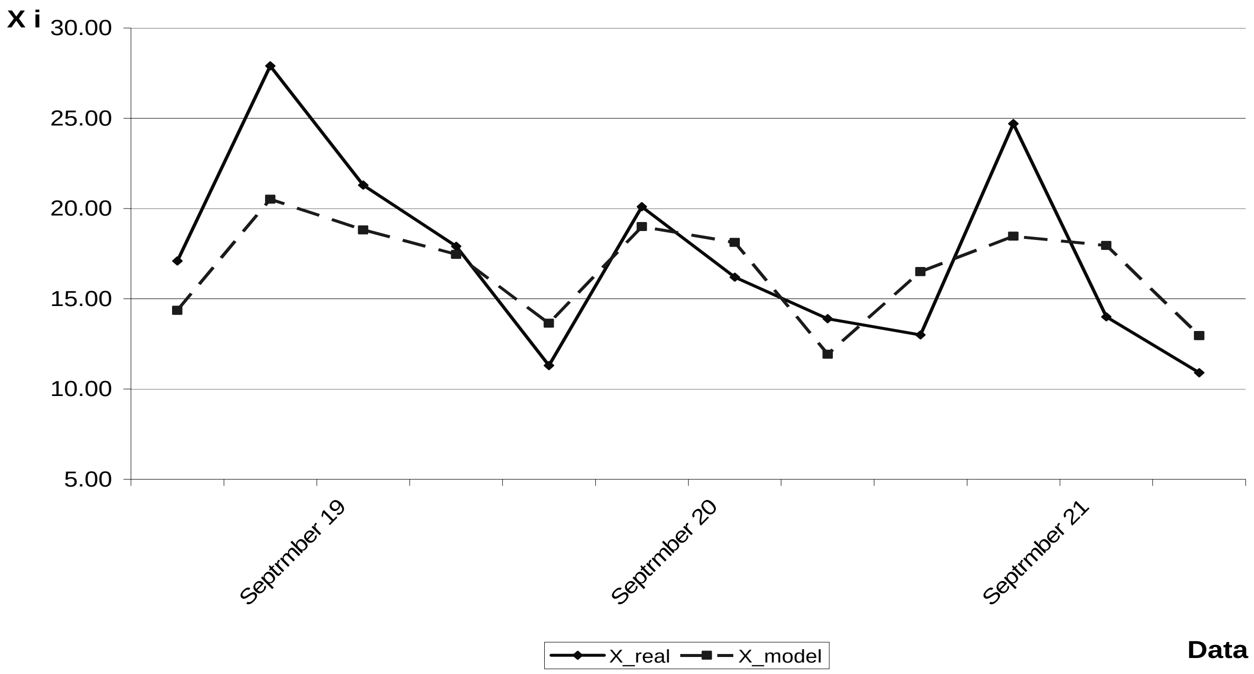

Figure 2. Real data of BGL and calculated by model (7)

Figure 6 shows that forecast using trend selection increases the model accuracy as compared to the forecast by combinatorial GMDH algorithm.

Conclusion

These examples show that the application method of trend selection leads to a reduction of the forecast error for the problem of forecasting the blood glucose level a patient with diabetes. The model obtained by Combinatorial GMDH algorithm (3) gives the average forecast error is 21.48%. Model obtained by calculating the deviation from the trend leads to the average prediction error - 17.35%.

It can be concluded that the selection of the regular component reduces the forecasting error. This reducing of accuracy is small so both models can be used to forecast the level of blood glucose level of a patient with diabetes.

References

1. Ivakhnenko A. G., Savchenko E.A. Inductive modeling in the task of diabetic treatment monitoring // Proceedings of international workshop on inductive modeling IWIM 2005. – IRTC, 2005. – P. 136-139 (In Russian).

2. Ivakhnenko A. G., Stepashko V.S.: Pomekhoustoichivost modelirovania (Noise-immunity of modeling). Kiev: Naukova Dumka, 216 p, 1985. (In Russian).

3. Ivakhnenko A.G., Savchenko E.A., Ivahnenko G.A., Gergely T. Usage of GMDH algorithms to renew missing data and forecast the blood glucose level under self-monitoring of diabetes // Problems of control and informatics. – 2002. – № 3. – P. 123-133. (In Russian).

4. Ivakhnenko A.G., Savchenko E.A. Investigation of Efficiency of Additional Determination Method of the Model Selection in the Modeling Problems by Application of the GMDH Algorithm // Journal of Automation and Information sciences. Begell House., Inc. Publishers.– 40, No 3.– 2008.– P.47-58.