ICMIEE18-299

Forecasting Demand by GMDH Predictor, a Case Study

Syed Misbah Uddin, Aminur Rahman, Emtiaz Uddin Ansari*
Department of Industrial and Production Engineering, Shah Jalal University of Science and Technology, Sylhet 3114, BANGLADESH

ABSTRACT

Forecasting demand is very important for manufacturing industry and also needed for all type of business and business suppliers for distribution of finish goods to the consumer on time. Forecasting with high accuracy is required to prevent wasting and system failure to fulfil market demand. This study is concerned with the determination of accurate models for forecasting cement demand. In this connection this paper presents results obtained by using a self-organizing model and compares them with those obtained by usual statistical techniques. A nonlinear modelling technique based on Group Method of Data Handling (GMDH) is considered here to derive forecasts. Various time series smoothing techniques such as exponential smoothing, double exponential smoothing, weightage moving average and moving average method are used for forecasting the demand. For this purpose, Monthly sales data of a typical cement ranging from January, 2007 to February, 2016 was collected. The mean absolute percentage error (MAPE) and mean sum square error (MSE) are also calculated for comparing the forecasting accuracy. The comparison of modelling results shows that the GMDH model perform better than other models based on terms of mean absolute percentage error (MAPE) and mean square error (MSE).

Keywords: Forecast, GMDH algorithm, Time series, MAPE, MSE.

1. Introduction

"Better predictions remain the foundation of all science. . . "[1]. Forecast accuracy has been a critical issue in the areas of financial, economic and scientific modeling, and has motivated the growth of a vast body of literature on the development and empirical application of forecasting models [2].More accurately forecasting demand would facilitate for assisting managerial, operational and tactical decision making. Therefore the selection of forecasting model is the important criteria that will influence to the forecasting accuracy [3].

GMDH method was originated in 1968 by Prof. Alexey G. Ivakhnenko in the Institute of Cybernetics in Kiev (USSR). This approach from the very beginning was a computer-based method so, a set of computer programs and algorithms were the primary practical results achieved at the base of the new theoretical principles. The method was quickly settled in the large number of scientific laboratories worldwide due to open code sharing. At that time code sharing was quite a physical action since the internet is at least 5 years younger than GMDH. Despite this fact the first investigation of GMDH outside the Soviet Union had been made soon by R.Shankar in 1972. Later on different GMDH variants were published by Japanese and Polish scientists.

The main idea of GMDH is the use of feed-forward networks based on short-term polynomial transfer functions whose coefficients are obtained using regression combined with emulation of the self-organizing activity behind NN structural learning [4]. To improve the performance of the GMDH algorithm, Barron gave a comprehensive overview of some early developments of network, and introduced the polynomial network training algorithm (PNETTR). Elder proposed Synthesis of Polynomial Network (ASPN) algorithm to improve the GMDH algorithm.

J.A.Muller and Frank Lemke developed and improved self-organizing data mining algorithms on the basis of the above results in 1990s. Further enhancements of the GMDH algorithm have been realized in the "KnowledgeMiner" software. The GMDH algorithm has gradually become an effective tool for modeling, forecasting, and decision support and pattern recognition of complex systems. There are processes for which it is needed to know their future or to analyze inter-relations. The GMDH algorithm was successfully used to deal with uncertainty, linear or nonlinearity of systems in a wide range of disciplines such as ecology, economy, medical diagnostics, signal processing, power plant, electric power industry and control systems [5-9]. The revised GMDH algorithms [10, 11]), have been introduced to model dynamic systems in flood forecast and petroleum resource prediction with some success.

The purpose of the study is to investigate how good is GMDH predictor as a forecasting tool by comparing the results of a self-organizing model with those obtained by usual statistical techniques. This paper organize as follows: section 2 describes the methodology applied for forecasting the demand. Data collection and analysis are presented in section 3. Section 4 provides the results and discussion while fifth section offers some conclusion.

2. Methodology

This is a case study research based on time series data of cement industry. The data used in this case study are monthly sales data of cement. The data span the period from January 2007 to February 2016. The dataset consists of 110 months' time series data. Data were analyzed by using various time series model such as moving average, weighted moving average, single

* Corresponding author. Tel.: +88-01770347090
E-mail addresses: misbah-ipe@sust.edu

exponential smoothing, double exponential smoothing and least square method of simple linear regression.

In this study, we use the value of α 0.3 and 0.5 for single exponential smoothing method. Simple exponential smoothing does not do well when there is a trend in the data. In such situations, several methods were devised under the name "double exponential smoothing" or "second-order exponential smoothing. The basic idea behind double exponential smoothing is to introduce a term to take into account the possibility of a series exhibiting some form of trend. This slope component is itself updated via exponential smoothing.One method sometimes referred to as "Holt-Winters double exponential smoothing are followed here. One of two*smoothing factor is* α which is called data smoothing factor and it's value, $0 < \alpha < 1$, and the other one β is the *trend smoothing factor*, $0 < \beta < 1$.

We also used the GMDH predictor version GMDH Data Science 3. 5. 9 to derive the forecast. Out of 110 data 58 months data are used for the training set and rest of the data are used for evaluation in checking set. In order to evaluate the forecasting accuracy of different techniques various central tendency measures as the loss function were also calculated with the help of following formula.

$$\text{MAD} = \frac{1}{\text{n}} \sum_{n=1}^{n} |(\text{ Actual} - \text{Forecast})|$$

n = the number of periods [12].

$$\text{MSE} = \frac{\sum_{k=0}^{n} \{\text{Actual} - \text{Forecast }\}^2}{\text{n}}$$

Where:

n = the number of periods [12].

$$\text{MAPE} = \frac{1}{\text{n}} \sum_{n=1}^{n} \frac{|\text{Actual} - \text{Forecast}|}{\text{Actual}} * 100\%$$

3. Data Collection and Analysis

Data Collection is a significant aspect of any type of research study. The data used in this case study are monthly sales data of cement. The data span the period from January 2007 to February 2016. The time series plot is given Fig. 1.

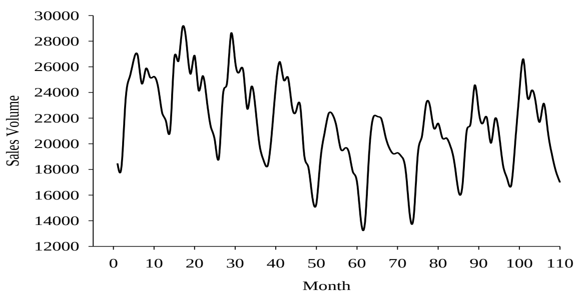

Fig.1 Monthly sales data (Jan 2007 to Feb 2016)

After collecting sales data GMDH algorithm and various statistical forecasting techniques were used to forecast. The mean absolute deviation (MAD), mean absolute percentage error (MAPE) and mean square error (MSE) were also calculated to assess forecasting performance of different models.

3.1 Analysis by GMDH algorithm

GMDH algorithm consists of set of steps that are described below:

Step 1: First *N* observations of regression-type data are taken. The collected load data are first normalized with respect to their individual base value in order to restrict the variation of data within the same level. Those normalized data are denoted by $(x_1, x_2, x_3, x_4, \dots\dots\dots\dots x_M)$ where M is the total number of input. The original data is separated into the training and test sets [14]. In this study total 110 data were separated into training (58) and test (52) sets. The 58 data is used for the estimation of the partial descriptions which describe the partial characteristics of the nonlinear system. The 52 data is used for organizing the complete description which describes the complete characteristic of the nonlinear system.

Step 2: Select $\binom{m}{2} = \text{m}(\text{m}-1)/2$ new input variables according to all possibilities of connection by each pair of inputs in the layer. Construct the regression polynomial for this layer by forming the quadratic expression which approximates the output *y* in equation (1).

Step 3: Identify the single best input variable out of these$\binom{m}{2}$ input variables, according to the value of mean square error (MSE). The input of variables that give the best results in the first layer, are allowed to form second layer candidate model of the equation (1). Set the new input $(x_1\, x_2 x_3\, x_4\ \dots\dots\dots\dots\ x_M)$ and $(M = M + 1)$ Models of the second layer are evaluated for compliance by using MSE, and again the input variables that give best results will proceed to form third layer candidate models. This procedure is carried out as long as the MSE for the test data set decrease compared with the value obtained at the previous one as shown in Fig. 2. After the best models of each layer have been selected, the output model is selected by the MSE. The model with the minimum value of the MSE is selected as the output model [15].

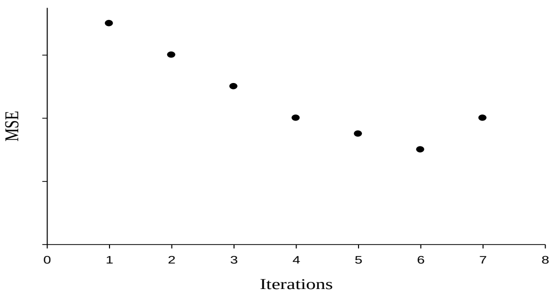

Fig.2 Stopping criteria of GMDH algorithm

3.2 Analysis by Statistical method

Various time series smoothing techniques such as exponential smoothing, double exponential smoothing, moving average and regression method were used for forecasting the load demand.Absolute deviations were also calculated. The mean absolute deviations (MADs) found from these calculations are listed in table 1.

Table 1MAD of different forecasting methods

Method	MAD
3 month Moving Average	2306
6 month Moving Average	2791
12 month Moving Average	2230
Weightage Moving Average	2056
Regression	2459
GMDH Method	704
Exponential α=0.3	2286
Exponential α= 0.5	2053
Double Exponential α= 0.3, β= 0.5	2861

From Table 1 it is seen that the value of MAD due to forecasting by GMDH algorithm is 704. On the other hand all the statistical method gives four digits MAD. The mean absolute percentage error (MAPE) and mean square error (MSE) were also calculated and reported in Table 2 and Table 3 respectively. It is observed that the GMDH forecast with only 4% MAPE and nearest value is 9% which is done by exponential smoothing technique(α= 0.5). Form Table 3 it is clear that the model with the minimum value of the MSE is the GMDH model.

Table 2MAPE of different forecasting methods

Method	MAPE
3 month Moving Average	11%
6 month Moving Average	14%
12 month Moving Average	11%
Weightage Moving Average	10%
Regression	12%
GMDH Method	4%
Exponential α=0.3	11%
Exponential α= 0.5	9%
Double Exponential α= 0.3, β= 0.5	13%

Table 3MSE of different forecasting methods

Method	MSE
3 Month Moving Average	7994519
6 Month Moving Average	10355301
12 Month Moving Average	7710194
Weightage Moving Average	6291543
Regression	9177720
GMDH Method	824882
Exponential α= 0.3	7619269
Exponential α= 0.5	6220179
Double Exponential α= 0.3, β= 0.5	11913465

4. Results and Discussion

After completing data analysis we have come out with some informative results. The calculated Mean absolute deviations (MADs) of forecasted data by different forecasting techniques are plotted in Fig. 3. It is seen that GMDH algorithm gives lowest value of MADwhich is best suit.

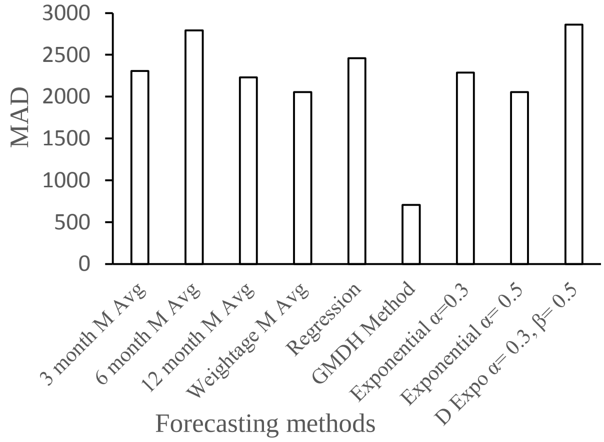

Fig.3 Comparison of MAD of different techniques

The mean absolute percentage error (MAPE) and mean square error (MSE) are plotted in Fig.4 and Fig.5 respectively. The comparison of modelling results shows that the GMDH model perform better than other models based on terms of mean absolute percentage error (MAPE) and mean square error (MSE).

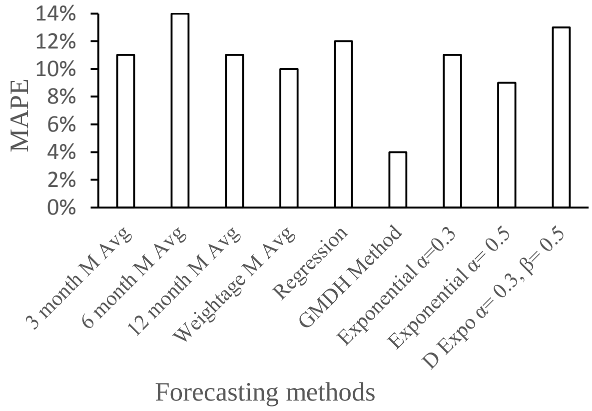

Fig.4 Comparison of MAPE of different techniques

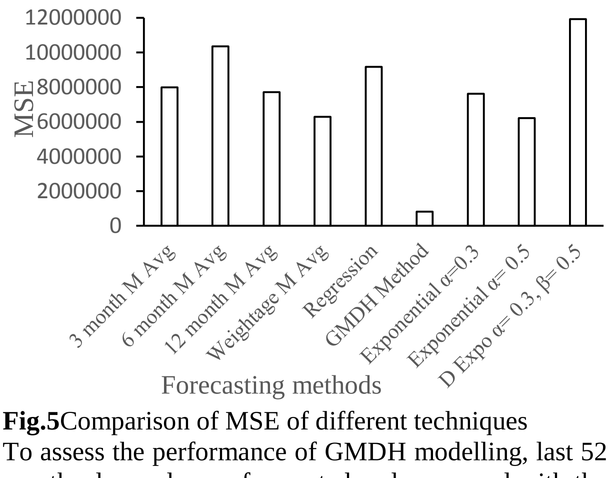

Fig.5 Comparison of MSE of different techniques

To assess the performance of GMDH modelling, last 52 months demand were forecasted and compared with the test set. The results of that model along with forecasting precision are shown in table 4. Normalize mean absolute error is found to be 4.65% whereasnormalize RMS is 6%. The fitting accuracy of GMDH model algorithm is also very good as the value of R^2 is 0.90.

Table 4 Summary results of GMDH modelling

Metrics	Output / Value
Post processed result	Model fit
Number of observations	52
Normalize mean absolute error	4.65%
Normalize root mean square error	6%
Standard deviation of residuals	5.8%
Coefficient of determination (Rˆ2)	0.90
Correlation coefficient	0.95

Our findings have several important implications. Useless input variables are eliminated and useful input variables are selected automatically, the structure parameters and the optimum GMDH architecture can be organized automatically. The case study on the cement time series data testing demonstrated that the GMDH model is robust in the forecasting of nonlinear time series.

5. Conclusion

This paper examined the forecasting accuracy of different statistical techniques as well as GMDH predictor. For that purposes ten years secondary sales data of a cement were collected. There was low seasonal variation in their sales. Demand forecasting was performed using extrapolative time series methods, such as exponential smoothing with level, trend, and seasonal components. Besides that moving average, weighted moving average and regression method were also used for forecasting the demand. A nonlinear self-organizing model based on Group Method of Data Handling (GMDH) was also applied here to derive forecasts.We applied the GMDH predictor version GMDH Data Science 3. 5. 9.

In order to evaluate the accuracy of prediction, various performance measures such as MAD. MAPE and MSE were calculated. It is found that there is no result near to the GMDH predictor. GMDH algorithm forecast with only 0.0367 or 4% error which is substantially more accurate than statistical method.

REFERENCES

[1] S.Makridakis, M. Hibon, The M3-Competition: results, conclusions and implications,*International Journal of Forecasting*,Vol. 16,pp451–476 (2000).

[2] J. Gooijer, R. Hyndman, 25 Years of Time Series Forecasting, *International Journal of Forecasting*, Vol. 22, pp 443– 473 (2006).

[3] R. Samsudin, P. Saad, A. Shabri, Hybridizing GMDH and Least quares SVM support vector machine for forecasting tourism demand, *IJRRAS*, Vol. 3, pp 274-279 (2010).

[4] S. J. Farlow, The GMDH Algorithm of Ivakhnenko, *The American Statistician*, Vol. 35, pp 210-215 (1981).

[5] A.G. Ivakheneko, G. A. Ivakheneko, A Review of Problems Solved by Algorithms of the GMDH, *Pattern Recognition and Image Analysis*, Vol. 5, pp 527-535 (1995).

[6] G. C. Onwubolu, P. Buryan, F. Lemke, Modeling Tool Wear in End-Miling Using Enhanced GMDHLearning Networks, *International Journal of Advance Manufacture Technology*, Vol.39, pp 1080–1092 (2008).

[7] T. Kondo, A. S. Pandya, A. S. Nagashino, GMDH-Type Neural Network Algorithm with a Feedback Loop for Structural Identification of RBF Neural Network, *International Journal of Knowledge-Based and Intelligent Engineering Systems*, Vol. 11, pp 157-168 (2007).

[8] V.Puig, M. Witczak, F. Nejjari, J. Quevedo, J. Korbicz, A GMDH Neural Network-Based Approach to Passive Robust Fault Detection Using a Constraint Satisfaction Backward Test, *Engineering Applications of Artificial Intelligence*, Vol 20, pp 886-897 (2007).

[9] F. Li, B. R. Upadhyaya,L. A. Coffey, "Model-Based Monitoring and Fault Diagnosis of Fossil Power Plant Process Units Using Group Method of Data Handling, *ISA Transactions*, Vol. 2, pp 213-219 (2009).

[10] T. Kondo, Nonlinear Pattern Identification by Multi-Layered GMDH-Type Neural Network Self-Selecting Optimum Neural Network Architecture. *International Conference on Neural Information Processing*,pp 882-891 (2007)

[11] F. Chang, Y. Hwang,A Self-Organization Algorithm for Real-Time Flood Forecast,*Hydrological processes*, Vol. 13, pp 123-138 (1999).

[12] P. K. Sahu, R. Kumar, Demand Forecasting for Sales of Milk Product (Paneer) in Chhattisgarh, *International Journal of Inventive Engineering and Sciences*, Vol. 1, pp 10-13 (2013).

[13] O. S. Ezennaya, O. E. Isaac, U. O. Okolie, Analysis of Nigeria's National Electricity Demand Forecast (2013-2030), *International Journal Of Scientific & Technology Research*, Vol. 3, pp 333-340 (2014).

[14] R. Samsudin,P. Saad, Combination of Forecasting Using Modified GMDH and Genetic Algorithm,*International Journal of Computer Information Systems and Industrial Management Applications*, Vol. 1, pp 170-176(2009).

[15] T. Kondo, J. Ueno, Revised GMDH-type Neural Network Algorithm with a Feedback Loop Identifying Sigmoid Function Neural Network, *International Journal of Innovative Computing, Information and Control*, Vol. 2(5), pp 985-996 (2006).