Stock Market Prediction using Inductive Models

Turki Y. Abdalla
Department of Computer Engineering
College of Engineering
University of Basrah

Abdulkareem Y. Abdalla
Department of Computer science
College of science
University of Basrah

Lamya A. Dexon
Department of Computer
science College of science
University of Basrah

ABSTRACT

Stock market Prediction is an example of a prediction problem which is challenging due to small sample sizes, high noise, and non-linearity. Neural networks(NN) have been frequently used in stock market prediction because of their ability to deal with uncertain, or insufficient data. Group Method of Data Handling (GMDH) is an inductive approach which attempts to overcome the limitations of neural networks based on the principle of self-organization. We have developed an algorithm based on the evolutionary manner of conventional GMDH NN to generate inductive model that can avoid some of GMDH problems like the exhaustive computations on candidate Adalines and the increasing number of Adalines in the following layers. The developed method is applied for stock market prediction**.** Simulation results are presented to demonstrate the capability of the proposed method for accurate prediction.

Key words

 stock market, inductive model , neural network, genetic algorithm, financial prediction

1. INTRODUCTION

The capability of induction is fundamental for human thinking. It is the next human ability that can be utilized in soft-computing, besides that of learning and generalization. The induction means gathering small pieces of information, combining it, using already collected information in the higher abstraction level to get complex overview of the studied object or process. Inductive modeling methods utilize the process of induction to construct models of studied systems [1]. The inductive methods have a similar concept to that of evolution introduced by Holland where a number of solutions are created and an external criterion plays the role of finding the fittest [2]. Inductive methods are objective, they can produce a self organization models On the other hand, Deductive methods are subjective since they are totally based on detailed instructions from the person performing the modeling as well as his/her idea about the object behavior [3].

Many attempts have been made to predict the behavior of stock markets, or other economic markets [4]. These attempts were encouraged by various evidences that economic markets behave in a chaotic manner [5]. Neural networks (NNs) have become very important method for finance applications because of their ability to deal with uncertain, fuzzy, or insufficient data which fluctuate rapidly in very short periods of time [6]. NNs differ from conventional techniques in that the analyst is not required to specify the nature of the relationships involved; the analyst simply identifies the inputs and the outputs [7,8].

Despite the small number of assumptions in NNs in comparison to statistical methods, they still require a significant amount of a priori information about the model's structure. Experts should decide on the quality and quantity of input arguments, the number of hidden layers and neurons as well as the form of their activation function [7].

2. GMDH NEURAL NETWORKS

GMDH neural network is an inductive approach which attempts to overcome the limitations of neural networks based on the principle of self-organization. An inductive approach is similar to neural networks but is unbounded in nature, where the independent variables of the system are shifted in a random way and activated so that the best match to the dependent variables is ultimately selected [8]. An idea behind GMDH algorithms which developed by Ivakhnenko (1966) is based on an evolution principle introduced by Holland which implies the generation and selection of the candidate neurons and an external criterion plays the role of finding the fittest [9].

The GMDH network is a kind of multilayered networks composed by a suitable combination of an Adalines which is illustrated in Figure 1
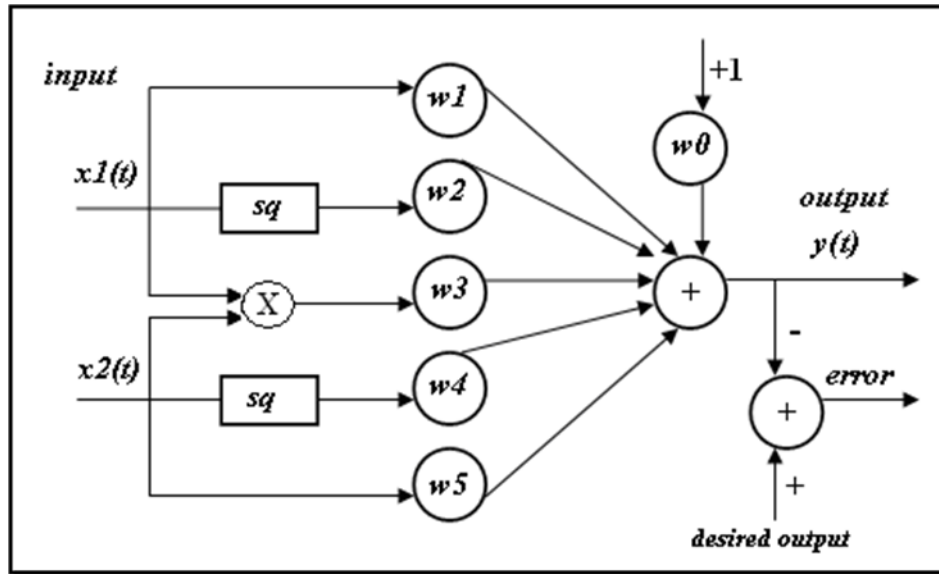
.

Figure 1: Structure of Adaline with a bi-quadratic polynomial function

The output variable y(t) is expressed as follows:

$$y(t) = w_1\, x_1(t) + w_2\, x_{12}(t) + w_3\, x_1(t)\, x_2(t)$$
$$+ w_4\, x_{22}(t) + w_5\, x_2(t) + w_o \qquad ..(1)$$

Where w_i are weights, and are determined by the least squares method and x_1, x_2 are the inputs to the unit.

Let $\mathbf{W} = [w_0\; w_1\; w_2\; w_3\; w_4\; w_5]^T$ and

$$\mathbf{X} = [1\; x_1\; x_1^2\; x_1 x_2\; x_2^2\; x_2]^T$$

The Widrow – Hoff delta rule for training W is as follows [9, 10]:

$$W_{k+1} = W_k + \beta\,(X_k\, /\, |X_k|^2)(y_d^{\,k} - W_k^{\,T}\, X_k) \qquad ..(2)$$

Where $y_d^{\,k}$ is the desired output at time k, β is the learning rate with the range (0,1) and $|X_k|^2$ is the square of length of input vector. The application of equation (2) causes W to be modified so that the difference between the desired and actual outputs will be reduced. The network structure is determined by the following steps [9, 11, 12]:

1- separate the experimental data set into a training data set and a selection data set;

2- create a layer of N Adalines based on the number of inputs. Every pair of inputs produces an Adaline. A combination C_2^i Adalines are created if there are i input variables;

3- use the training data to train all Adalines in the created layer and estimate the weighing coefficients for each Adaline with the least squares method;

4- input the selection data set to the network and obtain the error of all trained Adalines in the layer. Assign a threshold; keep the Adalines whose errors are below the threshold and use them to make the next layer;

5- If the smallest error of the current layer is larger than that of the previous layer or the current layer has only one Adaline, stop the training and trim the network.

6- test the performance of the trained network with evaluation data set. The evaluation data could be a combination of the training and selection data sets or a completely new data set.

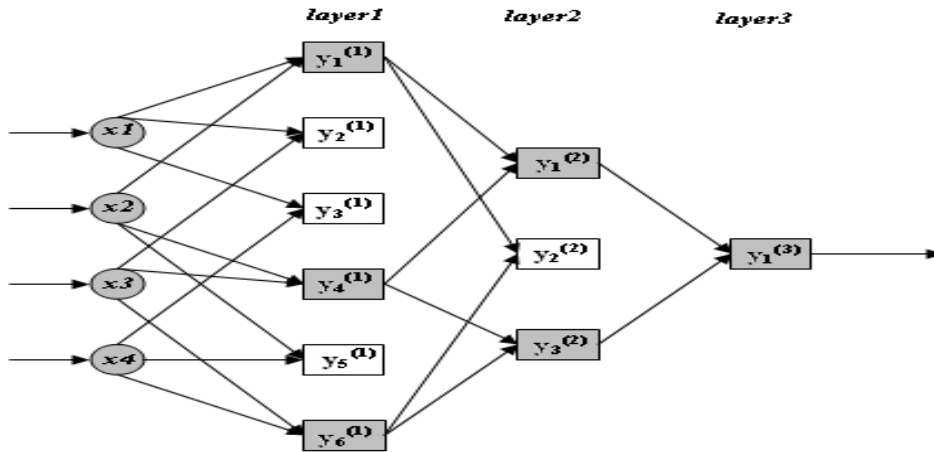The schematic diagram of a GMDH network with selection factor F=0.5 is shown in Figure2.

layer1　　　　　layer2　　　　layer3

Figure 2 : The structure of neural network grown by GMDH algorithm

The Adalines that were selected after trimming process at the layers are depicted in Figure 2 as the gray boxes. A resulting network is the three-layer network consisting of six Adalines and four input nodes

3. THE PROPOSED MODEL

An idea behind our algorithm is to select the neurons one-by-one and then add them to a binary network according to an external criterion . Only the neuron with the best selection error is added to an inductive network where the independent variables of the system are shifted in a random way and activated so that the best match to the dependent variable is ultimately selected. Following such a procedure there is a gradual increase of complexity and the optimum model is found which has a binary tree structure that is shown in Figure 3.
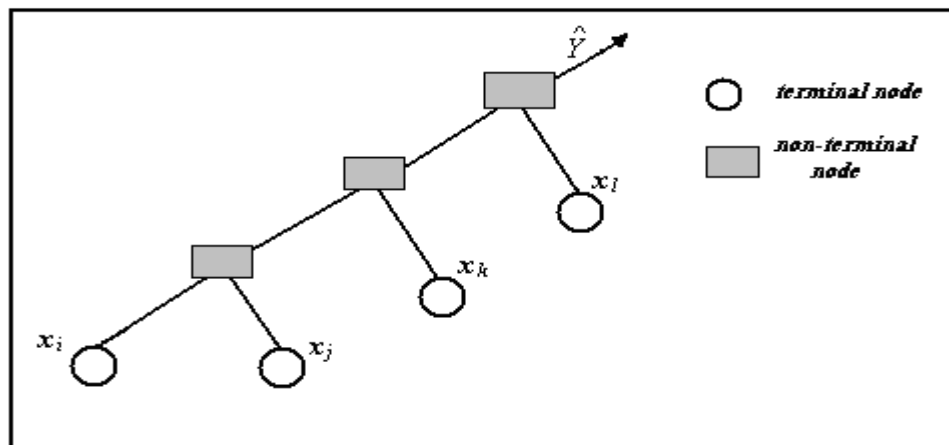

Figure 3: A binary tree with four terminals

\The method use Genetic Algorithm in the network construction to find the best one from a pool of individuals (chromosomes) that represents proposed binary tree networks. The chromosome which has the form that described in Figure 4 consists of **z** genes where **z** is the total number of candidate inputs. Each gene is an integer number (**i**) take values 0, 1,…, **z**.
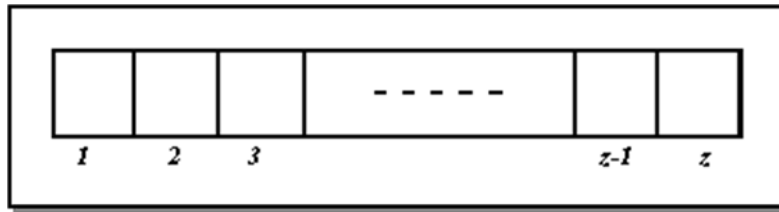
Figure 4: the form of the networks chromosome

The non zero integer i means the **i** [th] input variable is active, and the zero means **i** [th] input variable is not active.

The algorithm can be summarized as follows:

Step1: **initialization:** a population of **N** individuals (chromosomes) is randomly created. Each chromosome represents a proposed binary network. The chromosome has the form in Figure 4.

Step2: **N** networks are constructed corresponding to **N** chromosomes then trained using training data set by suitable training algorithm(delta rule or genetic algorithm) .

Step3: **evaluation:** a selection stage is performed by testing the generalization ability of all trained network to unseen data (selection data).

Step4: **Reproduction:** Genetic Algorithm operators of Rank selection, Uniform crossover, One-point mutation, evaluation and Replacement is repeatedly performed until a stopping condition is satisfied.

4. SIMULATION RESULTS

The value of a stock market price is established by analyzing the fundamental information associated with the market. The fundamental information consists of several input parameters which are used as input/output for our proposed prediction models. The input variables are:

1. Listed companies: number of listed companies in the market within a time period.

2. Exchange rates: the average of exchange rates between the local currency and US dollar within a period.

3. Market capitalization: Sum of the capitalizations of all listed companies in the market within a time period.

4. Value traded: the overall value of stocks that traded in the market within a time period.

5. Shares traded: sum of the shares (stocks) that traded in the market within a time period.

6. Days traded: number of the traded days in the stock market within a time period.

7. Turnover (TOR): is the result of division the value traded on the market capitalization in a time period multiplied by 100.

TOR = (value traded / capitalization) * 100

8. Capitalization average: is the result of the division of the Gross Domestic Product Average (GDPA) of the country to its market capitalization.

Capitalization average = GDPA / capitalization

The output variable is the price of the market's stock and it is computed as:

Stock price = value traded / shares traded.

A Saudi stock market is taken as a case study. The application data consists of the monthly averaged value for each of the input parameters described above. There are 141 trading months within the period from January 1996 to September 2007 taken from the website www.amf.org [13].

The input and the output data sets to the networks are preconditioned by normalizing the data. The purpose of normalizing the data is to modify the variable levels to a reasonable value. Normalizing data entails manipulating it to be in the range between 0 and 1. The networks are trained utilizing the normalized data sets. In our work data are first preprocessed by log function then normalization is done by dividing each data item on the greatest input value. The proposed inductive model is applied with real data to test the performance. A population of 30 binary members (proposed networks) is randomly initialized and the resulted models for two different schemes are described below:

4.1. Prediction using Inductive model with Delta-rule training algorithm [model A]

The network of the inductive model shown in Figure 5 is a final result of performing Genetic Algorithm to select the best network from a networks population. Genetic algorithm parameters are selected as 0.8 for crossover and 0.2 for mutation.
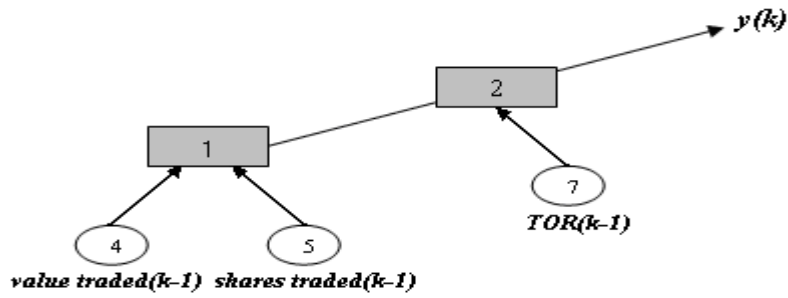
Figure 5: Trained inductive model A network

The trained network is a binary tree that consists of two Adalines trained with Delta Rule learning algorithm; first Adaline is trained for 2000 training cycles while the second is trained for 1500 training cycles only.

The adopted learning rate was 0.04 for both Adalines. The test results of the evaluation stage are described in Figure 6.
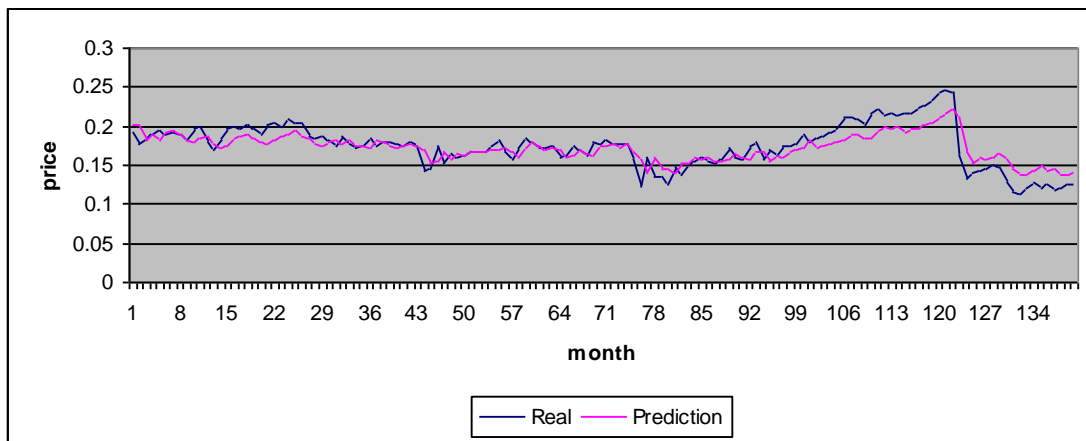

Figure 6: Inductive model A prediction results

4.2. Prediction using Inductive model with GA as a training algorithm [Model B]

Figure 7 shows a binary network that is the final result of performing Genetic Algorithm with parameters 0.8 for crossover and 0.1 for mutation to select the best

network. The five Adalines are trained using Genetic Algorithm to find the best weight coefficients of a 30 real member's population. Each Adaline is trained for 3000 genetic algorithm cycles with parameters of 0.6 for crossover and 0.2 for mutation. The results that are obtained from applying the whole 141 months data to the trained Inductive model B at the evaluation stage are given in Figure 8.
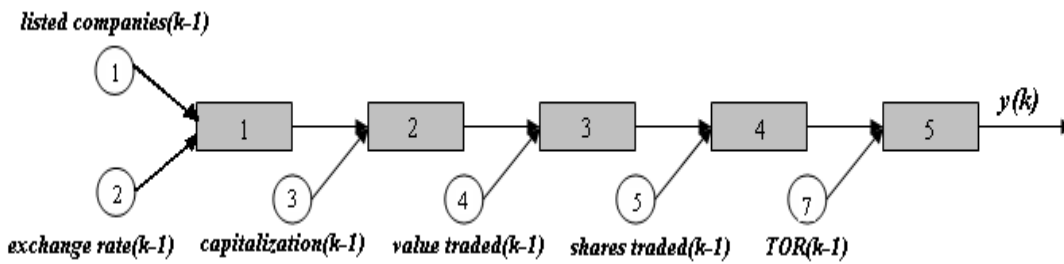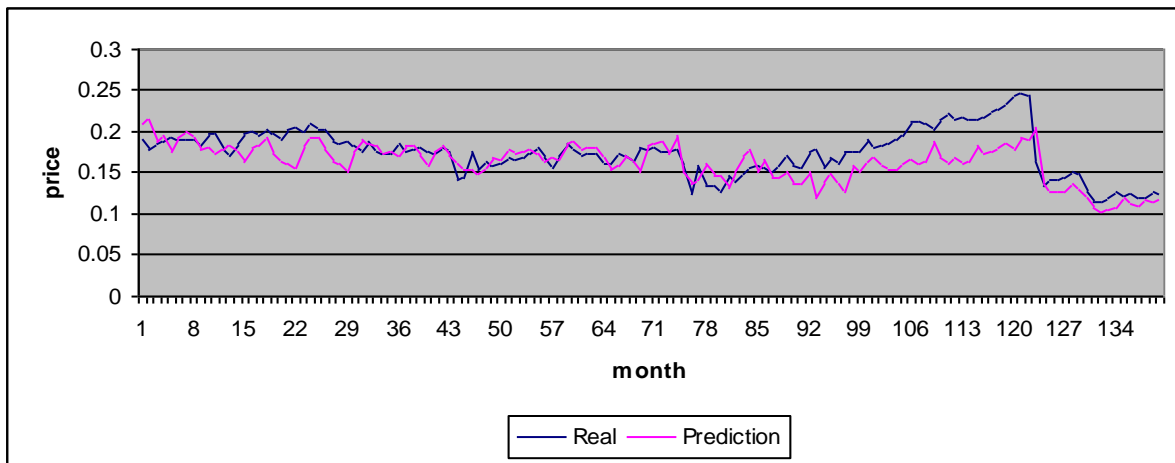
Figure 7: Trained inductive model B

Figure 8: Inductive model B prediction results

6 . CONCLUSION

The proposed Adaline based inductive model gives good prediction results as illustrated in Figure 7 and Figure 8; For comparison, the MSE errors at the evaluation stage are computed for the proposed inductive model and the conventional GMDH NN model as shown in Table 1. Inductive model A gives the best performance.

$$MSE = \frac{1}{N} \sum_{k=1}^{N} (y_d(k) - y(k))^2$$

Where: N is the number of month, $y_d(k)$ is the desired output and y (k) is the network output at month k.

Table 1: MSE of the proposed model for stock market prediction application.

No	Method	MSE
1	Biquadratic GMDH NN	0.0004955
2	Proposed Inductive Model A	0.0002404
3	Proposed Inductive Model B	0.0009048

5. REFRENCES

[1] P. Kordik, M. Snorek, M. Genyk-Berezovskyj, ' ' Hybrid Inductive Models: Deterministic Crowding Employed", 2004.

[2] H-S Park, B-J Park, H-K Kim, and S-K Oh, " Self-Organizing Polynomial Neural Networks Based on Genetically Optimized Multi-Layer Perceptron Architecture", International Journal of Control, Automation and Systems, vol. 2, no. 4, pp. 423-434, 2004.

[3] L. Anastasakis, and N. Mort, "The Development of Self-Organization Techniques in Modelling: A Review of the Group Method of Data Handling (GMDH)", 2001.

[4] K. Chakraborty, M. Kishan, K. M. Chlukuri, and R. Sanjay, "Forecasting the Behavior of Multivariate Time Series Using Neural Networks", Neural Networks, 5,96 1-970, 1992.

[5] M. E. Malliaris, "Modeling the Behavior of the S&P500 Index: A Neural Network Approach", Proc. 10th IEEE Conference on AI for Applications, 86-90, 1994.

[6] E. Schoeneburg, "Stock Price Prediction Using Neural Networks: A Project Report", Neurocomputing, vol. 2, 1990, pp. 17-27.

[7] N. W. Tan, "A Hybrid Financial Trading System Incorporating Chaos Theory, Statistical and Artificial Intelligence/Soft Computing Methods", Queensland Finance Conference 1999.

[8] G. R. Weckman, and R. Agarwala, "Identifying Relative Contribution of Selected Technical Indicators in Stock Market Prediction", 2003

[9] A. G. Ivakhnenko, and J-A. Muller, "Recent Developments of Self-Organising Modelling in Prediction and Analysis of Stock Market", 2000.

[10] M. Caudill, and C. Butler, "Understanding Neural Networks: Computer Explorations", 2000.

[11] V. Schetinin, J. Schult, and A. Brazhnikov, "Neural Network Techniques for visual mining clinical electroencephalograms", 2005.

[12] A. Sakaguchi, and T. Yamamoto, "A Study on System Identification Using GA and GMDM Network", IEEE 2003. (86) Arab Monetary Foundation, www.amf.org