ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEM FOR STOCK INDEX PREDICTION

Dinesh K. Sharma, Fayetteville State University, USA
profdksharma@gmail.com

H.S. Hota, Guru Ghasidas Central University, Bilaspur (C.G.), India
profhhota@gmail.com

ABSTRACT

In this paper, we explore the effectiveness and robustness of the Adaptive Neuro-Fuzzy Inference System (ANFIS) for time series data prediction in terms of number and types of fuzzy membership function. DOW30 index data was used to check the performance of the predictive model. Four different ANFIS models based on the type and number of fuzzy membership function are developed in order to see the effect of these two parameters. Empirical results show that the ANFIS model with three membership functions of type bell shape performs better than other ANFIS models at the training stage while an ANFIS model with two membership functions of type trapezoidal outperforms at the testing stage.

Keywords: Adaptive Neuro-Fuzzy Inference System (ANFIS), Fuzzy Membership Function (MF), Fuzzy Inference System (FIS).

INTRODUCTION

Forecasting in the financial scenario, especially in the stock market, is full of uncertainty which makes the stock prediction a very difficult task for investors and financial managers. Various factors are involved which can affect the forecasting and may cause a financial loss to the individual, financial institution, or industrial organization. Numerous traditional statistical techniques such as autoregressive integrated moving average (ARIMA) (Box & Jenkins, 1976), generalized autoregressive conditional heteroskedasticity (GARCH) (Franses & Ghysels, 1999) and smooth transition autoregressive (STAR) (Sarantis, 2001) have been used to predict stock price movement. However, these methods do not produce a better result and require some basic assumptions or continued review and refinement as the economic condition changes (Trippi and Turban, 1996). In order to provide better tools to predict stock price movement, a good and reliable system must be developed. Several authors have suggested Artificial Neural Network (ANN), Fuzzy Logic (FL), and Genetic Algorithm (GA) based systems for time series forecasting.

Over the past two decades, several financial researchers have applied ANN based techniques due to its ability of learning and pattern recognition. Studies show that ANN is a very powerful technique to develop an intelligent system for stock index forecasting. White (1988) used a simple ANN to perform time series analysis on the IBM common stock daily returns while Lin and Lin (1993) applied ANN to forecast the Dow Jones Industrial Average (DJIA). Trippi and DeSieno (1992) suggested technical analysis to investigate the effectiveness of a specific neural network trading system for S&P 500 index futures contracts. Choi et al. (1995) used ANN to predict the daily change in the S&P 500 index. Wang and Leu (1996) presented ARIMA based


Although Adaptive Network-Based Fuzzy Inference System (ANFIS) has been applied in several studies, few of these have contributed to research in the financial area (Boyacioglu and Avci, 2010). Atsalakis and Valavanis (2009b) develop a neuro-fuzzy adaptive control system to forecast the next day’s stock price trends of the ASE and the NYSE index. Boyacioglu and Avci (2010) investigated the predictability of stock market returns with ANFIS. Also, Esfahanipour & Aghamiri (2010) and Tan et al. (2011) used ANFIS for stock market analysis.

An ANN may have the problem of local minima and overlearning while on the other hand, integrating fuzzy logic with ANN may have problems due to deciding the universe of discourse and shape of membership function. Also, due to multiple inputs and 2-3 linguistic variables in each input, it is very difficult to generate a fuzzy inference system (FIS) for the model. Therefore, ANFIS may be the best alternative to capture the nonlinearity nature of stock market data. In this paper, we have used ANFIS for time series data predictions in terms of number and type of fuzzy membership function to stabilize the ANFIS models. The ANFIS models were tested on the DOW30 index data set to check the performance of the predictive model.

**DATA AND ANFIS TECHNIQUE**

The DOW30 daily price index data of about 12 years from Feb 1, 2000 to Jan 1, 2012 are used to train and test ANFIS models, large amounts of data sets are beneficial in terms of obtaining high accuracy from the ANN based model. The amount of training and testing samples must be divided in such a manner so that the ANN based model can learn the input/output pattern sufficiently. Therefore, the data set is divided into two parts-training and testing; training part consist of 2250 samples which is 75% of the total data set while the latest 750 samples (25%) are considered as the testing partition. The training ratio must be always higher than the testing ratio to stabilize the model properly. Training samples are used to build the ANFIS models while testing samples are used to validate the models. Model validation is a process in which input vector from input/output data sets that are unseen by the models are presented to the trained ANFIS models to see how well models are able to predict the corresponding output data.

ANFIS was first introduced by Jang (1993) by embedding the Fuzzy Inference System (FIS) into their framework of adaptive networks. ANFIS supports Sugeno FIS whose output is always crisp unlike Mamdani type FIS where output is always fuzzy. ANFIS uses the least-square and back propagation gradient descent methods along with a hybrid learning algorithm to tune membership function parameters and to build fuzzy if-then rules as the FIS with a single output. An adaptive network is a network structure consisting of a number of nodes connected through directional links with 5 layers (Jang et al., 1997) as below:
**Layer 1:** Node in this layer produces membership value, after receiving inputs.
**Layer 2:** This layer calculates firing strength of each rule by mathematical multiplication
**Layer 3:** This layer is a normalization layer which normalizes the strength of all rules.
**Layer 4:** This is the difuzzification layer. The output of each node in this layer is the product of the normalized firing strength and a first order polynomial.
**Layer 5:** The node in this layer computes the overall output as the summation of all incoming signals.

**ACCURACY MEASURING CRITERIA**

In time series prediction, it is a major question that which measuring criteria will reflect the effectiveness of the model in a proper way. Some very well-known error measures: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are widely used in this direction; the lower values of these measures are better in terms of efficiency of the model. Mathematical expressions of all these error measures are as follows:

\[
MAE = \frac{1}{N} \sum_{i=1}^{n} |A_i - P_i| \quad (1)
\]
\[
MAPE = 100 \times \frac{1}{N} \sum_{i=1}^{n} \frac{|A_i - P_i|}{A_i} \quad (2)
\]
\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (A_i - P_i)^2} \quad (3)
\]

Where \( A_i \) is the actual (Target) Next-Day-Close price, \( P_i \) is produced by the models and \( N \) is the total number of index data samples.

**EXPERIMENT WITH MATLAB**

Experimental work is carried out with MATLAB software (Mathworks Inc, USA) under Windows7 environment and Intel i5 processor; two tool boxes neural network and fuzzy logic were used to perform the work. MATLAB code is also written to simulate the work. In order to check the efficiency of the ANFIS model, four different ANFIS models based on the number and type (Shape) of fuzzy membership function (MF) are developed, because they may play a crucial role to a build robust predictive model. We have investigated four different membership functions and finally, bell shape and trapezoidal membership functions are considered to build ANFIS models. The details of these models based on the number and type of membership function are as follows:

- ANFIS-1: Two Bell shape membership functions.
- ANFIS-2: Three Bell shape membership functions.
- ANFIS-3: Two Trapezoidal membership functions.
- ANFIS-4: Three Trapezoidal membership functions.

An ANFIS model with four input parameters (Open, High, Low and Close) and with one output parameter Next-Day-Close with three bell shaped membership function is shown in Figure 1. The input variables (Open, High, Low and Close) are very important and most influence parameters for predictive output; hence, other variables are ignored from the data set. An initial membership function is designed with suitable universe of discourse as shown in Figure 1(a)
with three membership functions: Low, Medium and High, while membership functions after training is shown in Figure 1(b). Training of all the four ANFIS models are performed one by one by feeding 2,250 training samples for 100 epochs; training goal was meet with 100 epochs. During the training membership, functions are tuned by ANFIS in such a direction so that the model can produce more accuracy with the optimum set of rule. The designed architecture of ANFIS, in the case of three and two membership functions with four inputs and one output, are shown in Figure 2 (a) and 2 (b) respectively. After training the models, 16 and 81 number of rules is generated by ANFIS in case of two and three membership function respectively for each input. The number of membership function of input variables play a very important role, by an increasing number of membership, an ANFIS model will became more complex in terms of rule generated and will take more time to learn and infer the pattern. Hence, a fuzzy rule base must be optimized. It will be better if less number of rules in fuzzy rule base will produce better results. ANFIS-1 and ANFIS-3 are having 16 rules ($2^4$) as two membership functions are associated with each input in these cases while ANFIS-2 and ANFIS-4 are having 81 rules ($3^4$) as three membership functions are associated with each input in these cases.

(a)  
(b)  

Figure 1: (a) Initial bell shape membership function for Open and High input variables, (b) Final bell shape membership function for Open and High input variables

(a)  
(b)  

Figure 2: Architecture of ANFIS (a) With two membership functions (16 Rules), (b) With three membership functions (81 Rules).
RESULT ANALYSIS

After training all the ANFIS models one by one, results are obtained and models are stabilized. Then testing is performed with a testing data set consisting of 750 samples and model performances are evaluated. Predicted data, in case of all the four ANFIS models, is compared with actual index price and the prediction results are summarized in Table 1 in terms of three error measures explained in equations 1, 2 and 3. All the error measures: MAE, MAPE, and RMSE are least in case of training of ANFIS-2. MAE, MAPE and RMSE in this case are 81.113, 0.790 and 111.845 respectively while the opposite trends are there in case of testing of the models. ANFIS-3 is producing better results in case of testing; MAE, MAPE, and RMSE in this case are 32.795, 0.951 and 143.586 respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>MF Type</th>
<th>No. of MF</th>
<th>No. of Fuzzy Rules</th>
<th>Training MAE</th>
<th>Training MAPE</th>
<th>Training RMSE</th>
<th>Testing MAE</th>
<th>Testing MAPE</th>
<th>Testing RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANFIS-1</td>
<td>Bell Shape</td>
<td>2</td>
<td>16</td>
<td>85.339</td>
<td>0.833</td>
<td>119.626</td>
<td>34.152</td>
<td>1.0028</td>
<td>229.452</td>
</tr>
<tr>
<td>ANFIS-2</td>
<td>Bell Shape</td>
<td>3</td>
<td>81</td>
<td><strong>81.113</strong></td>
<td><strong>0.790</strong></td>
<td><strong>111.845</strong></td>
<td>39.304</td>
<td>1.2000</td>
<td>221.807</td>
</tr>
<tr>
<td>ANFIS-3</td>
<td>Trapezoidal</td>
<td>2</td>
<td>16</td>
<td>85.465</td>
<td>0.834</td>
<td>120.549</td>
<td><strong>32.795</strong></td>
<td><strong>0.951</strong></td>
<td><strong>143.586</strong></td>
</tr>
<tr>
<td>ANFIS-4</td>
<td>Trapezoidal</td>
<td>3</td>
<td>81</td>
<td>84.200</td>
<td>0.819</td>
<td>116.536</td>
<td>35.570</td>
<td>1.0225</td>
<td>264.603</td>
</tr>
</tbody>
</table>

At the training stage, ANFIS-2 is outperforming and then ANFIS-4 is producing better accuracy. On the other hand, at testing stage ANFIS-3 is performing better than all other models. Performance of the model at the testing stage is just opposite to that of the training stage, hence, ANFIS-2 is the best fitted model in case of training while ANFIS-3 is the best fitted model in case of testing. Accuracy obtained in case of testing is more reliable and important as compare to training because the prediction for unseen data is done at the testing stage.

This is also important that the model with only 16 rules in case of ANFIS-3 is performing better than an ANFIS model with 81 rules (ANFIS-4). Accuracy of the models is increasing in the case of less number of membership functions at the testing stage; also the trapezoidal membership function is performing better than the bell shape membership function. While the trends are just opposite at the training stage, accuracy of the models or increasing by an increasing number of membership functions. Also, model with bell shape membership function is performing better than the trapezoidal. ANFIS models output are also presented graphically as comparative graph in between actual and predicted next day close price in Figure 3 (a) and (b) for the best ANFIS.
model (ANFIS-3) at training and testing stages respectively, as it can be seen from these figures that the predicted index price is very close to the actual index price; however, due to the highly nonlinear nature of stock index data, there is a slight variation.

CONCLUDING REMARKS

Forecasting in the financial scenario especially in the stock market is full of uncertainty, which makes the stock prediction a very difficult task for investors and financial managers. Thus, we need proper predictions of the stock market related time series data. ANN is promising techniques used by many authors. However, due to the problem of local minima and over learning, it can be integrated with fuzzy logic. Fuzzy logic is basically used for approximation or reasoning while ANN is for training and adaption the pattern. By integrating these two techniques, we can be benefited with the features of both. In this study, we explored ANFIS neuro fuzzy techniques on the basis of a number and shape of the membership function. These ANFIS models are trained and then tested with the DOW30 data set. Empirical results show that an ANFIS model (ANFIS-3) with two membership functions of trapezoidal type produces the best accuracy with MAPE=0.951, MAE=32.795 and RMSE=143.586 with only 16 rules at the testing stage. High efficiency with less fuzzy rules may also beneficial in terms of complexity of the model. This study proves that a hybrid of ANN and Fuzzy Logic in the form of ANFIS may be the best alternative for the stock market prediction.

REFERENCES


(A complete list of references is available upon request from Dinesh Sharma at profdksharma@gmail.com)