Chapter 3

A Hybrid Approach to Forecast Stock Market Index

3.1 Introduction

In this chapter, we have used Adaptive Neuro Fuzzy Inference System (ANFIS) to forecast the stock market. ANFIS structure is the combination of fuzzy theory and artificial neural networks. The stock market is a complex, evolutionary and nonlinear dynamic system. The price variation of the stock market is chaotic, inherently noisy and non-stationary. The forecasting of the stock market is an attractive but still difficult activity in the modern business world. There are a number of factors either economic or non-economic which influence the behavior of the stock market.

Two major drawbacks are found in existing stock market prediction models (Wei et al. [92]):

- The rules produced by artificial intelligence forecasting models (neural networks (NN) and genetic algorithms) are complex and unintelligible.

- Statistic forecasting models, such as time series, are based on mathematical equations and require some basic assumptions for variables, which are not easily understandable by stock investors.

To overcome these drawbacks, we have used ANFIS combined with subtractive clustering to forecast the stock market. As we know, data dimensions are decreased with
the help of subtractive clustering. So firstly, subtractive clustering is used to class-
ify data set into useful linguistic values and then an ANFIS structure is used in the
prediction of stocks. ANFIS is the combination of fuzzy theory and artificial neural
networks to forecast the stocks.

The aim of this study is to investigate the predictability of stock markets, Bombay
Stock Exchange (BSE30), Hang Sang China Stock Exchange (HS), Japan Stock
Exchange (NIKKEI) and Taiwan Weighted Index (TWI).

3.2 Methodology

BSE30, HS, NIKKEI and TWI are the leading stocks in Asia. We evaluated the
proposed model with data taken from BSE30, HS, NIKKEI and TWI. The stepwise
description of the model is given as follows:

Step 1: Collection of data:
The data sample has been taken from BSE30, HS, NIKKEI and TWI index. Fore-
casting of future price fluctuations based on an examination of previous price fluc-
tuations applies to any tradable instrument, where the price is determined by the
forces of supply and demand. Price refers to any combination of the open, high,
low, or close for a given security over a specific time frame. The time frame could
be daily, weekly or monthly. Here in this study, daily stock prices with the four
variables (open, high, low, close) is used for prediction. Training data for BSE30
has been taken from Jan 1, 2011 to Oct 31, 2011 (207 observations) and testing
data from Nov 1, 2011 to Dec 31, 2011 (39 observations). Similarly, all the details
about training and testing data for HS, NIKKEI and TWI are given in Table 3.1.
To reduce the computational time, one-year data set has been used to evaluate the
proposed model.

Step 2: Clustering of data:
Data set is clustered by using a subtractive clustering (described in section 2.7.2)
method for making fuzzy rules. It is assumed that each data point in subtractive
clustering has a potential cluster center and calculates a measure of the likelihood
for each data point that would define the cluster center, based on the density of
Table 3.1: Training and Testing Data

<table>
<thead>
<tr>
<th>Stock Index</th>
<th>Training data</th>
<th>Testing data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>From</td>
<td>To</td>
</tr>
</tbody>
</table>

surrounding data points. Radii have a value between 0 and 1 and specify the size of the cluster in each of the data dimensions, assuming the data fall within a unit hyper box. Specifying a smaller cluster radius will usually yield smaller clusters in the data. The algorithm is given as follows:

1. Data point with the highest potential is selected as the first cluster center.

2. All data points in the vicinity of the first cluster center are removed (as determined by radii), to determine the next data cluster and its center location.

3. Iterations are done on the process until all of the data are within radii of a cluster center.

We have formed two, three and four clusters for each stock market with the help of MATLAB software. For Example, three clusters are formed for each data set by setting the parameters of clustering as follows:

- range of influence = 0.5,
- squash = 1.15,
- accept ratio = 0.3,
- reject ratio = 0.2.

**Step 3: Generation of forecasting model:**

To create an adaptive neural fuzzy network for training, we require a Fuzzy Interface System (FIS) (described in section 2.5). FIS is used to determine the initial parameters. To produce a FIS, we obtained linguistic intervals for input membership functions from Step 2. Here triangular membership function is used as input and
output membership function. Now to generate fuzzy ‘if-then’ rules, input membership functions are used in the ‘if’ condition part and output membership function in ‘then’ condition part. The general rule for BSE30 is given as follows:

\[ x(BSE30) = A_i, \ y(BSE30) = B_i, \ z(BSE30) = C_i \text{ then } f_i = p_i x + q_i y + r_i z + u_i, \]

where \( x(BSE30), y(BSE30) \) and \( z(BSE30) \) are linguistic variable, \( A_i, B_i \) and \( C_i \) are the linguistic values, \( f_i \) denotes the \( i \)-th output and \( p_i, q_i, r_i \) and \( u_i \) are the parameters. Similarly, we have obtained fuzzy ‘if-then’ rules for HS, TWI and NIKKEI stock markets.

Now the parameters of FIS membership functions from training data are optimized by utilizing the combination of least-squares and the back propagation gradient descent method. This study sets 100 epochs as the training stopping criterion. Once the stopping criterion is reached, the FIS parameters for ANFIS are determined.

**Step 4:** The proposed model for forecasting the future price:

After obtaining the optimal membership functions, the data set is trained along with optimal fuzzy rules and membership functions to forecast the future price \( S(t + 1) \) in the testing data sets.

**Step 5:** Evaluation of forecasting performance:

The performance of proposed model is evaluated with the help of correlation coefficient, which is given as:

\[
 r = \frac{\sum_{i=1}^{N} (actual(t_i) - \overline{actual(t)}) (forecast(t_i) - \overline{forecast(t)})}{\sqrt{\sum_{i=1}^{N} (actual(t_i) - \overline{actual(t)})^2 \sum_{i=1}^{N} (forecast(t_i) - \overline{forecast(t)})^2}}, \quad (3.2.1)
\]

where \( actual(t_i) \) denotes real value of data, \( \overline{actual(t)} \) is mean of real value of data, \( forecast(t_i) \) is predicted value and \( \overline{forecast(t)} \) is the mean of predicted value.

Also, to compare the performance of proposed model with the existing models, root mean square error (RMSE) is chosen as evaluation criterion defined as:

\[
 RMSE = \sqrt{\frac{\sum_{i=1}^{N} |actual(t_i) - forecast(t_i)|^2}{N}}, \quad (3.2.2)
\]

where \( actual(t_i) \) refers to the actual value of \( i^{th} \) data point, \( forecast(t_i) \) refers to the predicted value of \( i^{th} \) data point and \( N \) is the total number of data entries.
## 3.3 Experimental Results

The proposed model is implemented by using the daily stock price of BSE30, HS, NIKKEI and TWI stock exchanges. The four major attributes: open, low, high, and close from the stock market are used in this model. The forecasting variable is the next day’s closing price. Firstly, we cluster the data by using subtractive clustering technique. Two, three and four clusters are formed for each stock index. In the second phase, FIS is built for each set of clusters using the corresponding training data. Finally, at the testing stage, data are clustered and forecasting is done for each set of clusters.

From Table 3.2, we observed that the correlation coefficient of actual and predicted values are positive for every stock index. TWI gives a high value of correlation coefficient i.e., 0.998 when number of clusters are three for predicted value and BSE, HS, NIKKEI have higher correlation coefficient values 0.991, 0.996, 0.995 respectively for two clusters. Positives values of Correlation coefficient show that actual and predicted values are very closely related to each other, hence giving very good forecasting results.

A comparison of the proposed model with ARIMA ([16]) and Chen’s ([25]) models in terms of RMSE is shown in Table 3.3 and results have shown that the proposed model gives relatively better forecast than some existing models in the literature. From Table 3.3, we can see that RMSE value for TWI index is lower when number of clusters are three and higher for two clusters. Similarly, BSE, HS and NIKKEI have minimum RMSE when no. of clusters are two. Hence for the final prediction, we have selected three clusters for TWI and two for BSE, HS and NIKKEI. Also, from the Figure 3.1, we can see that the TWI perform best in case of three clusters and BSE, NIKKEI and HS give the best performance in case of two clusters. Figures 3.2-3.5 show the forecasting and actual closing values in the testing phase for all the stock markets are very close, which shows the effectiveness of the proposed model.
Table 3.2: Correlation coefficient for different no. of clusters for each stock market

<table>
<thead>
<tr>
<th></th>
<th>TWI</th>
<th>NIKKEI</th>
<th>HS</th>
<th>BSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Cluster 2</td>
<td>0.993</td>
<td>0.995</td>
<td>0.996</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>Cluster 3</td>
<td>0.998</td>
<td>0.765</td>
<td>0.987</td>
</tr>
<tr>
<td></td>
<td>Cluster 4</td>
<td>0.996</td>
<td>0.992</td>
<td>0.990</td>
</tr>
<tr>
<td>Actual</td>
<td>Cluster 2</td>
<td>0.995</td>
<td>0.992</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>Cluster 3</td>
<td>0.994</td>
<td>0.823</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>Cluster 4</td>
<td>0.997</td>
<td>0.994</td>
<td>0.987</td>
</tr>
</tbody>
</table>

Table 3.3: Comparison of Proposed model with existing models in terms of RMSE

<table>
<thead>
<tr>
<th></th>
<th>TWI</th>
<th>NIKKEI</th>
<th>HS</th>
<th>BSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed model Cluster 2</td>
<td>27</td>
<td>23</td>
<td>82</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Cluster 3</td>
<td>25</td>
<td>59</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Cluster 4</td>
<td>26</td>
<td>24</td>
<td>84</td>
</tr>
<tr>
<td>Existing Models Chen ([25])</td>
<td>27</td>
<td>35</td>
<td>89</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>ARIMA ([16])</td>
<td>32</td>
<td>48</td>
<td>93</td>
</tr>
</tbody>
</table>

3.4 Conclusion

The main role of this study was to forecast the stock market with the help of ANFIS. We clustered the data by using subtractive clustering; then we generated a fuzzy inference system by using two, three and four clusters and then ANFIS was used to optimize the parameters of FIS. ANFIS uses the combination of gradient descent and least square method to optimize the parameters of the membership functions. Hence, the proposed model adopts the past trends of the stock prices and gives better forecasting when the trends are monotonic. We can see from Figures 3.2-3.5 the prediction is much better when the trends are monotonic. We evaluated the performance of proposed model over four stock markets and results were compared with some existing models. Result reveals that the proposed model gives relatively better forecast as compared to the listed models. Therefore, proposed hybrid fuzzy time series approach will be a good choice for forecasting of the stock market index.
Figure 3.1: RMSE values of forecasting results for each set of clusters

Figure 3.2: Forecasting results of the BSE index
Figure 3.3: Forecasting results of the TW index

Figure 3.4: Forecasting results of the HS index
Figure 3.5: Forecasting results of the NIKKEI index