Chapter 5

Minimal Variability OWA Operator Combining ANFIS, FCM and Subtractive Clustering to Predict BSE Index

5.1 Introduction

In stock market prediction, high-order time series models which use past multiple days of stock prices as forecasting factors are more sensible to give a superior portfolio for stock investors than one-order time series model which uses only previous one day’s stock data. Nevertheless, in prediction procedure, high order time series data is very difficult to handle. Since it is hard to assign a proper weight to every period of previous stock value, so reduction of data dimensions without losing the original information and suggestion of a comprehensive prediction model based on stock prices of nonlinear relationships is difficult. Also, numerical time series models have been developed to address prediction issues for stock markets, but these models can not tackle the non-linear relationships among the stock prices.

In this chapter, a new forecasting model is proposed extending the model presented in the previous chapter. Past forecasting models cannot process data base with higher dimensions easily, because complexity will increase multi-fold with the growth
of data dimensions.

To address all issues listed above, this chapter incorporates Order Weighted Averaging (OWA) and Adaptive Neuro Fuzzy Inference System (ANFIS) technique. OWA is used to reduce the computational complexity of high dimensional data. After the reduction of dimensions, following two models are proposed to produce understandable rules for investors and stock forecasting.

- Model 1: ANFIS with fuzzy c-means clustering (FCM).
- Model 2: ANFIS with subtractive clustering.

The proposed models are verified through an empirical analysis of the stock data sets, collected from Bombay Stock Exchange (BSE30) and the results are compared with some existing models in the literature. Results have shown that proposed model gives relatively better forecast than the existing models.

5.2 Basic Terminology and Tools

In this section, we have stated some results, which have been used in the subsequent work of this chapter.

5.2.1 Order Weighted Averaging

The OWA first introduced by Yagar [96], has generated much interest among researchers. In recent years, many related studies have been conducted. Fuller and Majlender [45] used Lagrange multipliers to solve constrained optimization problem and determined the optimal weighing vector. Fuller and Majlender [46], employed the Kuhn-Tucker second order sufficiency conditions to optimize and derive OWA weights.

- Yagar’s OWA

Yagar [96], proposed an OWA to get optimal weights of the attributes based on the rank of these weighing vectors after processing aggregation.

An OWA operator of dimension $n$ is a mapping $f : R^n \rightarrow R$ that has an associated weighting vector $W = [w_1, w_2, ..., w_n]^T$ with the following properties:
$w_i \in [0, 1]$ for $i \in I = \{1, 2, 3, ..., n\}$ and $\sum_{i \in I} w_i = 1$ such that

$$f(a_1, a_2, ..., a_n) = \sum_{i \in I} w_i b_i, \quad (5.2.1)$$

where $b_i$ is the $i$th largest element in the collection of the aggregated objects $a_1, ..., a_n$.

In [96], Yagar introduced two important characterizing measures associated with the weighing vector $W$, i.e., measure of orness and measure of dispersion, which are defined as follows:

$$\text{orness}(W) = (1/n - 1) \sum_{i=1}^{n} ((n - i) * w_i) = \alpha, \quad (5.2.2)$$

so orness$(W) = \alpha$ is a situation parameter and $0 \leq \alpha \leq 1$.

$$\text{disp}(W) = -\sum_{i=1}^{n} w_i \ln w_i. \quad (5.2.3)$$

The orness measure has the following property: if $W = \{w_1, w_2, ..., w_n\}$ is the weight vector of an OWA with orness$(W) = \alpha$, then $W' = \{w_n, w_{n-1}, ..., w_1\}$ is the reverse order of $W$, orness$W' = 1 - \alpha$.

- **Fuller and Majlender’s OWA**

Fuller and Majlender [46] transformed Yagar’s OWA equation to a polynomial equation by using Kuhn-Tucker second order sufficiency conditions. According to their approach, the associated weight vectors can be obtained as follows:

The interval $(0, 1)$ is partitioned by the following equation

$$(0, 1) = \bigcup_{r=2}^{n-1} J_{r,n} \cup J_{1,n} \cup \bigcup_{s=2}^{n-1} J_{1,s}, \quad (5.2.4)$$
where

\[
J_{r,n} = \left( 1 - \frac{2n + r - 2}{3(n-1)}, 1 - \frac{2n + r - 3}{3(n-1)} \right), \quad r = 2, 3, ..., n - 1, \quad (5.2.5a)
\]

\[
J_{1,n} = \left( 1 - \frac{2n - 1}{3(n-1)}, 1 - \frac{n - 2}{3(n-1)} \right), \quad (5.2.5b)
\]

\[
J_{1,s} = \left[ 1 - \frac{s - 1}{3(n-1)}, 1 - \frac{s - 2}{3(n-1)} \right), \quad s = 2, ..., n - 1. \quad (5.2.5c)
\]

When \( \alpha \in J_{r,s} \) then

\[
W^* = (0, ..., 0, w_r^*, ..., w_s^*, 0, ..., 0)^T, \quad (5.2.6)
\]

where

\[
w_j = 0 \quad \text{if} \quad j \notin I_{(r,s)}, \quad (5.2.7a)
\]

\[
w_r^* = \frac{2(2s + r - 2) - 6(n - 1)(1 - \alpha)}{(s - r + 1)(s - r + 2)}, \quad (5.2.7b)
\]

\[
w_s^* = \frac{6(n - 1)(1 - \alpha) - 2(s + 2r - 4)}{(s - r + 1)(s - r + 2)}, \quad (5.2.7c)
\]

\[
w_j^* = \frac{s - j}{s - r} w_r + \frac{j - r}{s - r} w_s \quad \text{if} \quad j \in I_{(r+1,s-1)}, \quad (5.2.7d)
\]

where \( I_{(r+1,s-1)} = r + 1, ..., s - 1 \).

If \( r = 1 \) and \( s = n \) then, \( \alpha \in J_{1,n} \), and

\[
W^* = (w_1^*, ..., w_n^*)^T, \quad (5.2.8)
\]
5.3 Proposed Model

In the recent past, Artificial Neural Network (ANN) and Genetic Algorithms (GA) have been applied in fuzzy time series (huarng and Yu [54], Zhang [105]) modelling. However, black box and the forecasting rules are quite complex to understood, so they are not very helpful to investors to buy and sell stocks. These days ANFIS is widely used for forecasting stocks. ANFIS can produce 'if-then' rules to framework the quantitative aspects of human knowledge. The rules are simple, hence useful for investors. ANFIS can process data without any underlying stationary or non-stationary assumptions, which are necessary for Autoregressive Moving Average Model (ARMA) (Box [16]), Autoregressive Conditional Heteroskedasticity (ARCH) (Engle [40]) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) (Bollerslev [15]). Moreover, it can also describe non-linear relationships which are lacking in the above statistical models (jilani and Burney[60]).

OWA operator is used to reduce the dimensions of stock data, results in the reduction of computational complexity, which otherwise is a potential problem in forecasting models employing multiple variables or periods of stock prices (Fuller and Majlender [45]).

A linear equation is employed to fuse recent periods of prices with the given weights into a single forecasting factor as input for ANFIS. In the fusion process, each attribute is multiplied by its corresponding weight and the weighted attributes are summarized as one aggregated forecasting factor. Because of linear calculations of OWA, data dimensions can be reduced with less computational complexity.
In a practical stock market, investors make their short term decisions based on recent stock fluctuations. Therefore the OWA operator is proposed to properly weigh short and long term market fluctuations. Applying OWA in forecasting models can widely improve the forecasting accuracy (Chang et al. [29]). Fuller and Majlender [45] have shown that weight based techniques provide an efficient method to increase the performance of time series models. The overall schematic flowchart of the proposed model is shown in Figure 5.1.

5.3.1 Steps of proposed algorithm

The algorithm of the proposed model is given as follows:

**Step 1:** Identification of the number of attributes for forecasting the future price:
The latest three periods of stock price, $S(t - 2)$, $S(t - 1)$ and $S(t)$ are used to forecast the future price $S(t + 1)$. So the number of attributes identified are three.

**Step 2:** Calculation of the OWA weights:
In this step, OWA weights are calculated by the equations (5.2.4-5.2.8). (Several $\alpha$ values, ranging from 0.1 to 0.5 are utilized to produce different sets of OWA weights.)

$\alpha = 1 - \alpha$ if $\alpha > 0.5$

Here each $\alpha$ value represents one set of influence degrees for the recent three periods of stock prices given in Table 5.1.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_1$</td>
<td>0.0</td>
<td>0.0333</td>
<td>0.1333</td>
<td>0.2333</td>
<td>0.3333</td>
</tr>
<tr>
<td>$W_2$</td>
<td>0.2</td>
<td>0.3333</td>
<td>0.3333</td>
<td>0.3333</td>
<td>0.3333</td>
</tr>
<tr>
<td>$W_3$</td>
<td>0.8</td>
<td>0.6333</td>
<td>0.5333</td>
<td>0.4333</td>
<td>0.3333</td>
</tr>
</tbody>
</table>

**Step 3:** Calculation of the aggregated value:
In this step, the aggregated value by one linear equation is calculated, which is given
as:

\[ A(t) = W_1 \times S(t) + W_2 \times S(t-1) + W_3 \times S(t-2). \]  (5.3.1)

where \( W_1, W_2 \) and \( W_3 \) denote different degree of influence for the three recent periods of stock prices \( S(t), S(t-1) \) and \( S(t-2) \). \( A(t) \) denotes the aggregated value of stocks with corresponding degrees of influence.

**Step 4: Clustering of data:**

After calculating the aggregated value, data clusters are formed using two different methods which are described as follows:

- **Subtractive Clustering**

  As we explained in the section 2.7.2 that every data point in subtractive clustering has a potential cluster center. Based on the density of surrounding data points a measure of the likelihood for each data point has been calculated that would define the cluster center. The value of Radii varies between 0 and 1, and also defines the cluster size in each of the data dimensions. We set the parameters as follows:

  - range of influence = 0.47,
  - squash = 1.25,
  - accept ratio = 0.5,
  - reject ratio = 0.15.

- **Fuzzy C-Means Clustering**

  FCM (section 2.7.1) is an iterative process. The starting fuzzy partitioning matrix is created and initial fuzzy cluster centers are computed. In every iteration, the distance between cluster centers and data points is minimized and then cluster centers and membership values are updated. The process stops automatically, when the specified number of iterations are attained or when the improvement between two consecutive iterative objective functions is below the minimum amount assigned. The parameters of FCM clustering are given below:
– maximum number of iterations = 100,
– exponent for the partition matrix = 2.0,
– minimum amount of improvement = 0.0067.

Three clusters are formed for each year’s data set in both the clustering techniques. These clusters are used as linguistic labels.

**Step 5:** Identification of the type of membership function for input and output variables:
Since $A(t)$ is the only aggregated value which is used as input variable. For this input, subtractive and FCM clustering generates three Linguistic intervals (explained in Step 4). Both the clustering methods use gaussian membership function as input membership function. To develop an output linguistic variable only one type of membership function is used corresponding to three linguistic intervals generated by subtractive and FCM clustering. Now fuzzy 'if-then' rules are generated by using Takagi-Sugeno fuzzy model with three inputs and one output, the description of generated fuzzy rules are given as follows:

If $x(A(t)) = A_i$ then $f_i(S(t+1)) = p_i x + r_i$,

where $x(A(t))$ denotes linguistic variable, $A_i$ gives the linguistic value, $f_i$ announces the $i^{th}$ output and $p_i$, $r_i$ are the parameters ($i = 1, 2, 3$).

**Step 6:** Generation of fuzzy inference system:
Three linguistic intervals (low, medium, high) for input are obtained by subtractive and FCM clustering. Membership functions for input and output are set in Step 5. The input linguistic intervals are used in the 'if' condition part and output membership function is in the 'then' condition part. The rules generated by fuzzy inference system are given as follows:

Rule 1. if $x(a_j) = A_{low}$ then $f_{low}(t+1) = p_{low} x + r_{low}$,

Rule 2. if $x(a_j) = A_{mid}$ then $f_{mid}(t+1) = p_{mid} x + r_{mid}$,

Rule 3. if $x(a_j) = A_{high}$ then $f_{high}(t+1) = p_{high} x + r_{high}$,

where $A_i$ denotes linguistic values, $x(a_j)$ denotes the linguistic variables, $f_i(t+1)$ is the $i^{th}$ output value and $p_i$, $r_i$ are the parameters.
Step 7: Training of parameters:
After the formation of fuzzy ’if-then’ rules, the combination of back propagation gradient-descent and least-square method is used in the training process of proposed models. The training data sets are used to find the more efficient parameters for the fuzzy inference system to get more accurate results. To improve the input parameters, ANFIS used the gradient-descent method. For output parameters, ANFIS used least-square method.

For example, rule 1 parameters for the year 2012 given by subtractive clustering and FCM are shown in Table 5.2 as follows:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Subtractive</th>
<th>FCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_{low} )</td>
<td>1.004</td>
<td>1.15</td>
</tr>
<tr>
<td>( r_{low} )</td>
<td>-71.83</td>
<td>-88.02</td>
</tr>
</tbody>
</table>

Step 8: Forecasting the future price:
After obtaining the optimal membership functions for both subtractive and FCM clustering methods, the data set is trained along with optimal fuzzy rules and membership functions in the target testing data set for the prediction of future price \( S(t + 1) \).

Step 9: Evaluation of prediction performance:
To compare the performance of proposed model with the existing models, root mean square error (RMSE) is chosen as evaluation criterion, defined as:

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

(5.3.2)

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.
Table 5.3: Performance comparison of different models in terms of RMSE

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RW [48]</td>
<td>179</td>
<td>248</td>
<td>384</td>
<td>277</td>
<td>182</td>
<td>231</td>
<td>181</td>
</tr>
<tr>
<td>Chen’s model [25]</td>
<td>140</td>
<td>164</td>
<td>291</td>
<td>217</td>
<td>124</td>
<td>181</td>
<td>121</td>
</tr>
<tr>
<td>ANFIS [56]</td>
<td>120</td>
<td>136</td>
<td>178</td>
<td>172</td>
<td>102</td>
<td>155</td>
<td>92</td>
</tr>
<tr>
<td>Cheng et al. [29] α=0.5</td>
<td>138</td>
<td>156</td>
<td>188</td>
<td>177</td>
<td>113</td>
<td>167</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>164</td>
<td>180</td>
<td>348</td>
<td>257</td>
<td>146</td>
<td>212</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>188</td>
<td>201</td>
<td>395</td>
<td>297</td>
<td>168</td>
<td>243</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>212</td>
<td>223</td>
<td>446</td>
<td>336</td>
<td>189</td>
<td>276</td>
<td>184</td>
</tr>
<tr>
<td></td>
<td>235</td>
<td>245</td>
<td>497</td>
<td>376</td>
<td>210</td>
<td>307</td>
<td>203</td>
</tr>
<tr>
<td></td>
<td>258</td>
<td>270</td>
<td>549</td>
<td>415</td>
<td>231</td>
<td>344</td>
<td>222</td>
</tr>
<tr>
<td>Cheng et al. [29] α=0.6</td>
<td>140</td>
<td>164</td>
<td>292</td>
<td>216</td>
<td>124</td>
<td>191</td>
<td>121</td>
</tr>
<tr>
<td>Cheng et al. [29] α=0.7</td>
<td>140</td>
<td>164</td>
<td>292</td>
<td>216</td>
<td>124</td>
<td>191</td>
<td>121</td>
</tr>
<tr>
<td>Cheng et al. [29] α=0.8</td>
<td>140</td>
<td>164</td>
<td>292</td>
<td>216</td>
<td>124</td>
<td>191</td>
<td>121</td>
</tr>
<tr>
<td>Cheng et al. [29] α=0.9</td>
<td>140</td>
<td>164</td>
<td>292</td>
<td>216</td>
<td>124</td>
<td>191</td>
<td>121</td>
</tr>
<tr>
<td>Cheng et al. [29] α=1.0</td>
<td>140</td>
<td>164</td>
<td>292</td>
<td>216</td>
<td>124</td>
<td>191</td>
<td>121</td>
</tr>
</tbody>
</table>

Proposed model 1 α=0.1 | 36   | 43   | 77   | 56   | 33   | 45   | 35   |
| α=0.2               | 68   | 80   | 143  | 105  | 62   | 87   | 61   |
| α=0.3               | 90   | 105  | 192  | 139  | 81   | 119  | 80   |
| α=0.4               | 115  | 134  | 241  | 176  | 102  | 147  | 100  |
| α=0.5               | 140  | 164  | 292  | 216  | 124  | 191  | 121  |

Proposed model 2 α=0.1 | 37   | 45   | 77   | 56   | 34   | 45   | 33   |
| α=0.2               | 68   | 83   | 158  | 105  | 62   | 84   | 64   |
| α=0.3               | 91   | 105  | 190  | 140  | 81   | 113  | 79   |
| α=0.4               | 117  | 134  | 244  | 177  | 105  | 146  | 100  |
| α=0.5               | 140  | 164  | 292  | 217  | 124  | 181  | 121  |

5.4 Experimental results

Data is taken from the Bombay Stock Exchange index BSE30, for the period of 2006 to 2012. BSE-30 is a free-float weighted Indian stock market index consisting of 30 financially sound and well-established companies. Each year of BSE30 is used as one complete verification data set. There are almost 240 to 260 trading days in each year. So, we have used the first 200 days as training data set and the rest as testing data set. Three rules are formed for each year.

We predicted the stock market based on the last three days historic values, i.e. the effect of longer observation periods had been studied. Table 5.1 shows that the
highest weight is given to the third day, and second highest to the second day and lowest to the first day. It is observed that the proposed models i.e. model 1 and model 2 give minimum RMSE value when $\alpha = 0.1$ which is shown in Figures 5.2 and Figure 5.3 respectively.

The experimental results have shown that for every model there are seven testing data sets each corresponding to a particular year, so total 35 forecasting performance records for different values of $\alpha$ ranging from 0.1 to 0.5 are shown in Table 5.3. In the same table a comparison in terms of RMSE of proposed models with Random Walk (RW) [48], Chen’s [25], ANFIS [56] and Cheng et al. [29] models are presented. Also, a comparative analysis of the forecasted RMSE obtained from the proposed models and existing models is shown in Figure 5.4 and Figure 5.5. From these Figures, it is observed that the RW model gives the maximum value of RMSE and proposed models give minimum RMSE value. It shows that the average performance of the proposed models is better than the compared models. Corresponding to the value $\alpha = 0.1$ proposed models give the best performance among the listed models with minimum RMSE value (Table 5.3).

From Figure 5.4 and Figure 5.5, we can see that the RMSE is very low in the case of the proposed models as compared to the model given by Chen [25]. Chen [25] has predicted university enrollments using fuzzy time series. The fuzzy logic system is easy to design but also contains some critical problems. As systems complexity increases, it becomes more difficult to find out the correct set of rules. Fuzzy logic uses the heuristic procedure in antecedent processing, rule formation and for defuzzification. But the heuristic procedure does not provide satisfactory solutions and fuzzy logic cannot give rules that would meet a pre-assigned accuracy. So to overcome all these drawbacks, fuzzy logic has been combined with ANN to give more accurate results. Besides, proposed models have used clustering techniques to form clusters. Clustering results for the year 2006 for model 1 and model 2 are shown in Figure 5.7 and Figure 5.8 respectively. Cluster centers are represented by black dots. Clusters can extract hidden structures in the data and help to form easy and more exact rules in a fuzzy inference system and help in the reduction of error.

Comparison of the proposed model 1 and model 2 is given in Figure 5.6, which
shows that model 1 will give better forecasting results than model 2 since it has a lesser value of RMSE. For $\alpha = 0.1$, the fuzzy membership functions for each year of model 1 are shown from Figures 5.9 to 5.15 and for model 2 from Figures 5.16 to 5.22.

5.5 Conclusion

Improvement of the time series forecasting accuracy has been an important aspect yet often difficult task for decision makers on many fronts. Since existing forecasting models have many drawbacks, in this chapter, we have developed a new time series model combining minimal variability OWA operator to aggregate high dimensional data into a single useful forecasting factor. ANFIS combined with fuzzy c-means and subtractive clustering was then used for forecasting stock prices. From the experimental results, it has been shown that the proposed models outperform some listed models. Further, the results of these models can be very useful for stock investors and decision makers. The proposed models generate only three linguistic rules, substantially reducing the computational complexity. It is also observed that the BSE30 index depends on last three days indices, which is contrary with TAIWAN index, which depends only on the previous day index as reported by Cheng et al. [29]. Hence the BSE30 index has longer observation period, compared to a small country index such as TAIWAN. The results have revealed that one can predict the stock market with more accuracy by looking at a longer past data. Due to time zone as well as delay on consequence, all stock market indices depends upon other countries stock market behavior.

Results have also shown that proposed model 1 gives lower RMSE values than model 2. So model one is a relatively better choice to predict the stock market behavior.

It is further observed that the proposed methods have some advantages, such as the power of decision making can be extended with the use of minimal variability OWA operator, as they can adjust the weights according to the situation of the decision maker. Since FCM is used in the fuzzification step, some problems caused
by the partition of discourse of the universe are removed. Also, there is no need
to use difficult matrix operations or complex fuzzy group relationship tables since
fuzzy relationships are defined by artificial neural networks. Therefore, proposed
hybrid fuzzy time series approach will be a good choice for forecasting of the stock
market index.
Figure 5.1: Proposed model sequence
Figure 5.2: Performance comparison of proposed model 1 for different $\alpha$ values in terms of RMSE

Figure 5.3: Performance comparison of proposed model 2 for different $\alpha$ values in terms of RMSE
Figure 5.4: Performance comparison of proposed model 1 with different models in terms of RMSE

Figure 5.5: Performance comparison of proposed model 2 with different models in terms of RMSE
Figure 5.6: Performance comparison of proposed model 1 and model 2 in terms of RMSE

Figure 5.7: FCM cluster center for the year 2006
Figure 5.8: Subtractive clustering cluster center for the year 2006

Figure 5.9: Model 1 final input membership function for the year 2006
Figure 5.10: Model 1 final input membership function for the year 2007

Figure 5.11: Model 1 final input membership function for the year 2008
Figure 5.12: Model 1 final input membership function for the year 2009

Figure 5.13: Model 1 final input membership function for the year 2010
1.5 1.6 1.7 1.8 1.9 2 2.1
\times 10^4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Data
Degree of membership

Figure 5.14: Model 1 final input membership function for the year 2011

1.55 1.6 1.65 1.7 1.75 1.8 1.85 1.9 1.95
\times 10^4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Data
Degree of membership

Figure 5.15: Model 1 final input membership function for the year 2012
Figure 5.16: Model 2 final input membership function for the year 2006

Figure 5.17: Model 2 final input membership function for the year 2007
Figure 5.18: Model 2 final input membership function for the year 2008

Figure 5.19: Model 2 final input membership function for the year 2009
Figure 5.20: Model 2 final input membership function for the year 2010

Figure 5.21: Model 2 final input membership function for the year 2011
Figure 5.22: Model 2 final input membership function for the year 2012