Applying Fuzzy C-Means and Artificial Neural Networks into a High-Order Fuzzy Time Series Prediction Model

Mu-Yen Chen, and Hsiu-Sen Chiang
Department of Information Management, National Taichung University of Science and Technology, Taichung, Taiwan

Abstract - A novel high-order fuzzy time series model for stock price forecasting is presented based on the fuzzy c-means (FCM) discretization method and artificial neural networks (ANN). In the proposed model, the FCM discretization method obtained reliable interval lengths. In addition, the fuzzy relation matrix was obtained from ANN, mooring the need for complex and time-consuming matrix operations. The proposed model was validated using experimental datasets from authentic university enrollment data. Empirical results indicate the proposed model outperforms existing methods for forecasting time series in terms of root mean squared errors (RMSE).

Keywords: Fuzzy time-series, fuzzy c-means, artificial neural networks

1 Introduction

The 2007–2010 global financial crisis involved housing asset bubbles, credit booms, predatory lending, incorrect pricing of risk and the collapse of the shadow banking system. From January to August 2010, the US FDIC (Federal Deposit Insurance Corporation) (2015) announced 111 U.S. bank failures, and an additional 231 failures occurred over the next four years [1]. The crisis has had a significant negative impact on the United States and the European Union. From 2010 to 2012, four euro-zone countries (Greece, Ireland, Portugal and Cyprus) suffered continuous negative growth, leaving national governments unable to meet national debt payments [2]. Recently, Business Insider (2015) has also claimed that “the euro crisis is entering a new, highly dangerous phase, and once again Greece finds itself at the centre” [3].

Time series forecasting is used to predict future performance in many fields, facilitating the preparation for and mitigation of potential crises including economic shocks, power outages and droughts. Such forecasts are currently made using methods including regression analysis, moving average, autoregressive conditional heteroscedastic (ARCH) models [4], Generalized ARCH (GARCH) [5]. However, these statistical models are highly reliant on historical data which must follow a Gaussian distribution to optimize forecasting performance. Moreover, traditional time series forecasting models cannot deal with fuzziness or uncertain datasets because they lack an accurate measurement of certain data, or have difficulty obtaining actual measured values [6]. These issues are commonly addressed using fuzzy time series which can be applied to linguistic value datasets to produce accurate forecasting results. Fuzzy time series models have been used successfully many times to forecast nonlinear datasets in such widely varying applications as course enrollment [7], power loading [8], wind speed [9] and stock market performance [10, 11, 12, 13].

Following the most recent global financial crisis and European debt crisis, many researchers have urged the development of novel time series models for predicting economic events, such as stock market crashes, housing bubbles and credit booms. Currently, financial forecasting relies mainly on mathematical and statistical methods [14, 15, 16], and time series models [17]. Many theories and techniques for financial forecasting have been used in fundamental and technical analysis. This research applies fuzzy c-means (FCM) and artificial neural networks (ANN) to the construction of a high-order fuzzy time-series model for use with university enrollment datasets. Future research will aim to solve time series problems for financial predictions of stock market performance.

2 Literature Reviews

2.1 Fuzzy Time-Series Definitions

Song and Chissom [18, 19, 20] first proposed the application of fuzzy theory in time series in 1993 to deal with problems involving uncertain linguistic information. Fuzzy time series are designed using fuzzy sets [21, 22] and can thus overcome the disadvantage of traditional time series which can only deal with real numbers. Recently, Song and Chissom [18, 19, 20] have presented additional definitions for fuzzy time series as follows [20].

Definition 1. A fuzzy sets $A(t)$ in the universe of discourse $U$, where $U = \{u_1, u_2, \ldots, u_n\}$ can be represented as follows:

$$A(t) = f_{A(T)}(u_1) + f_{A(T)}(u_2) + \ldots + f_{A(T)}(u_n)$$

where $f_A$ is the membership function of the fuzzy set $A(t)$ and $f_{A(T)}: U \rightarrow [0, 1]$, $f_{A(T)}(u_i)$ denotes the membership degree of $u_i$ in the fuzzy sets $A(t)$, $1 \leq i \leq n$.

Definition 2. Let $A(t), A(t-1), A(t-2), \ldots, A(t-n)$ be the fuzzy sets in a time series. Assume that $A(t)$ is caused by $A(t -$
1), $A(t-2), \ldots$ and $A(t-n)$. Then the $n$th-order fuzzy logic relationship can be represented as follows:

$$A(t-n), A(t-n+1), A(t-n+2), \ldots, A(t-1) \rightarrow A(t)$$

**Definition 3.** Let $A(t-1) = A_i$ and $A(t) = A_j$, where $A_i$ and $A_j$ are fuzzy sets. The fuzzy logical relationship (FLR) can be denoted by $A_i \rightarrow A_j$, where $A_i$ is called the left-hand side (LHS) and $A_j$ is the right-hand side (RHS) of this relationship.

### 2.2 Fuzzy Time-Series Models

Over the past decade, several modifications to fuzzy time-series models have been proposed and can be categorized into high-order models [23, 24, 25] and artificial intelligence approaches. Because the first-order fuzzy time series model is established only using a simple fuzzy set structure, it has difficulty dealing with complex fuzzy logical relationships [26, 27]. For this reason, Own and Yu (2005) proposed a novel high-order fuzzy time series model to overcome complex time series issues [28]. Li and Cheng (2007) presented a deterministic high-order fuzzy time series model to forecast enrollment at the University of Alabama [29]. Their proposed forecasting model outperformed existing conventional models. Singh (2009) used many different orders as forecasting parameters and employed a w-step fuzzy predictor to test against datasets for University of Alabama student enrollment and Lahi crop production [30]. The following year, Li et al. (2010) developed a deterministic high-order fuzzy time series model is established only using a simple fuzzy set structure, it has difficulty dealing with complex fuzzy logical relationships [28, 29]. Their proposed forecasting model outperformed existing conventional models. Singh (2009) used many different orders as forecasting parameters and employed a w-step fuzzy predictor to test against datasets for University of Alabama student enrollment and Lahi crop production [30].

For the artificial intelligence approach, Hsieh et al. (2011) integrated wavelet transformations and recurrent neural networks (RNN) to forecast international stock markets [31]. The same results were computed by RNN with Chandra and Zhang (2012) [32], and Smith and Jin (2014) [33]. Chen and Kao (2013) integrated particle swarm optimization and support vector machine techniques into their fuzzy time series model. Experimental results showed their proposed method outperformed existing conventional methods for stock market forecasts [34]. Recently, the same results were obtained that the support vector machine technique integrated with the proposed fuzzy time series models from Ruan et al. (2013) [35], Yang et al. (2015)[36] and particle swarm optimization approach with Pulido et al. (2014) [37], Singh and Borah (2014) [38], and Lin et al. (2015) [39].

### 2.3 Fuzzy C-Means Clustering Method

Bezdek’s (1981) fuzzy C-means (FCM) algorithm [40] is the most popular classical fuzzy clustering method. FCM tries to divide the dataset by minimizing the fuzzy least-square error objective function within the group with respect to the fuzzy membership $u_{ij}$ and center $v_i$:

$$J_{\beta}(X,U,V) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^\beta d^2(x_i;v_j)$$

wherein $\beta > 1$ for adjusting the noise in the dataset is the fuzziness index, $n$ is the number of feature vectors $x_i$, $c > 2$ is the number of clusters in the dataset, and $d(x_i; v_j)$ is the measure of similarity between a datum and a center. In addition, the $J_{\beta}$ minimizes the constraints as follows:

$$0 \leq u_{ij} \leq 1 \quad \forall i,j,$$

$$0 < \sum_{i=1}^{c} u_{ij} \leq n \quad \forall j,$$

$$\sum_{i=1}^{c} u_{ij} = 1 \quad \forall j,$$

generating a minimized pseudo-iterative algorithm known as the FCM algorithm. Each center $v_i$ and the membership degree $u_{ij}$ are updated according to the following expressions.

$$v_i = \frac{\sum_{j=1}^{c} u_{ij}^\beta x_i}{\sum_{j=1}^{c} u_{ij}^\beta}$$

$$u_{ij} = \frac{\left( \frac{d(x_i;v_j)}{d(x_i;v_i)} \right)^{\beta-1}}{\sum_{j=1}^{c} \left( \frac{d(x_i;v_j)}{d(x_i;v_i)} \right)^{\beta-1}}$$

where $j$ is a variable on the feature space, i.e. $j = 1, 2, \ldots, m$.

### 3 Research Methodology

The overall flowchart of the proposed model is shown in Fig. 1. The novel fuzzy time-series prediction model is built through two steps: pre-processing and optimization partitioning.

#### 3.1 Pre-processing Phase

**3.1.1. Data Source and Variable Selection**

A single simple variable, “Actual”, is used as the input variable to the proposed model to forecast student enrollment at the University of Alabama. Enrollment data from 1971 to 1992 are summarized in Table 1.
Table 1. Enrollment data of University of Alabama

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>13055</td>
</tr>
<tr>
<td>1972</td>
<td>13563</td>
</tr>
<tr>
<td>1973</td>
<td>13867</td>
</tr>
<tr>
<td>1974</td>
<td>14696</td>
</tr>
<tr>
<td>1975</td>
<td>15460</td>
</tr>
<tr>
<td>1976</td>
<td>15311</td>
</tr>
<tr>
<td>1977</td>
<td>15603</td>
</tr>
<tr>
<td>1978</td>
<td>15861</td>
</tr>
<tr>
<td>1979</td>
<td>16807</td>
</tr>
<tr>
<td>1980</td>
<td>16919</td>
</tr>
<tr>
<td>1981</td>
<td>16388</td>
</tr>
<tr>
<td>1982</td>
<td>15433</td>
</tr>
<tr>
<td>1983</td>
<td>15497</td>
</tr>
<tr>
<td>1984</td>
<td>15145</td>
</tr>
<tr>
<td>1985</td>
<td>15163</td>
</tr>
<tr>
<td>1986</td>
<td>15984</td>
</tr>
<tr>
<td>1987</td>
<td>16859</td>
</tr>
<tr>
<td>1988</td>
<td>18150</td>
</tr>
<tr>
<td>1989</td>
<td>18970</td>
</tr>
<tr>
<td>1990</td>
<td>19328</td>
</tr>
<tr>
<td>1991</td>
<td>19337</td>
</tr>
<tr>
<td>1992</td>
<td>18876</td>
</tr>
</tbody>
</table>

3.1.2. Data Splitting

Most studies verify forecasting performance using the data splitting method wherein the experimental dataset is split into two subsets, one for training and one for testing. However, in this research, data splitting is not needed because the forecasting performance will be evaluated for each year in the enrollment data.

3.1.3. Define the universe of discourse, U

In the traditional fuzzy time series model, we first select two numbers $D_{\text{min}}$ and $D_{\text{max}}$ as the respective minimum and maximum values in the target dataset. We can thus determine the universe of discourse $U$ as $[D_{\text{min}} - D_1, D_{\text{max}} + D_2]$, where $D_1$ and $D_2$ are two appropriate positive numbers. For example, based on enrollment data from 1971 to 1992, the $U$ of the enrollment data from the University of Alabama are 13000 and 19400 according to a minimum enrollment of 13055 and a maximum enrollment of 19337, where $D_1 = 55$ and $D_2 = 63$. However, the FCM discretization method was applied to obtain reliable interval lengths, thus no measuring methods were needed to determine a suitable interval length for the discrete fuzzy sets [41].

3.1.4. Fuzzy Time-Series Model

Many previous studies followed Miller (1994) in establishing a universe of discourse $U$ into seven intervals of equal lengths (i.e., $u_1, u_2, u_3, u_4, u_5, u_6, u_7$) because that human short-term memory function is seven, plus or minus two [42]. In this research, we also adopted the Miller’s (1994) approach to establish the fuzzy sets $A_1, A_2, ..., A_7$ of the universe of discourse $U$.

3.2 Optimization Partitioning Phase

3.2.1. FCM-based discretization partitioning

In this research, we adopted FCM to identify the optimization intervals based on Cheng’s [43] model, which can adjust the linguistic interval and allow the observations to be more reliably fuzzified into the optimal linguistic values. This step includes two sub-steps:

Step 1. Applying time series $T$ into $c$ fuzzy clusters

Assume a time series $T$ of $u$ attribute with $v$ observations. In this step, the appropriate fuzzy clustering method selects a cluster time series $T$ into $c$ ($2 \leq c \leq n$) cluster. By applying fuzzy clustering methods to the partitioning and fuzzifying process, each fuzzy cluster is used to predict the impact of fitness on the accuracy rate. In this step, FCM is selected because it is the most popular fuzzy clustering method. However, the optimal number of clusters remains an open and subjective question. Thus, following Miller (1994), seven clusters are used here to identify the cluster that corresponds to the limits of human perception.

As shown in Eq.(2), this step applies FCM to the time series $T$, and each clustering center is denoted as $v_{ij}$ (i = 1,2, ..., c and j = 1,2, ..., m). In advance, the Euclidean distance was used to measure the distance of each cluster iteratively. As shown in Eq.(3), the memberships $u_{it}$ (t = 1,2, ..., n) of each
cluster center was computed when the cluster centers are shown to be stable.

**Step 2. Fuzzifying the time series T(t) as fuzzy time series A(t)**

As described in Step 1, the fuzzy time series \( \{A(t) = [u_1, u_2, \ldots, u_c]\} \) could be caused by the time series \( T(t) \). In addition, this approach can easily determine whether \( A(t) \) belongs to cluster \( C_i \) when the maximum membership of \( A(t) \) occurs in cluster \( C_i \).

Thus, using \( C_i \) as the center of \( C_i \), each cluster is ordered according to its main attribute value rankings. The clusters are used to define the orderly linguistic variable \( L_r \) \((r = 1, 2, \ldots, c)\). For instance, assume there are three clusters with centers of 40, 120 and 80. Their centers are used to arrange them separately as \( C_1 \), \( C_3 \), and \( C_2 \), and define their corresponding linguistic variables as \( L_1 \), \( L_3 \), and \( L_2 \).

3.2.2. Artificial Neural Networks Optimization

This step includes two sub-steps used to verify our proposed model:

**Step 1. Use artificial neural networks to construct the FLRs**

When the time series model is complete, the fuzzy relationship between the different time orders can be deduced. The first-order FLR is always built from the previous fuzzy set \( A(t-1) \rightarrow A(t) \). For example, in Table 2, \( A(13867) \rightarrow A(14696) \) is a relationship, then a fuzzy relationship \( L_2 \rightarrow L_3 \) is used to substitute \( A(13867) \) and \( A(14696) \) with \( L_2 \) and \( L_3 \). Finally, the sequence of the FLRs is as follows: \( L_3 \rightarrow L_4, L_4 \rightarrow L_4, L_4 \rightarrow L_5, L_5 \rightarrow L_6, L_6 \rightarrow L_6 \) as shown in Table 2.

**Table 2. Linguistic values of enrollment data**

<table>
<thead>
<tr>
<th>Year</th>
<th>Enrollment data</th>
<th>Linguistic values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973</td>
<td>13867</td>
<td>( L_2 )</td>
</tr>
<tr>
<td>1974</td>
<td>14696</td>
<td>( L_3 )</td>
</tr>
<tr>
<td>1975</td>
<td>15460</td>
<td>( L_4 )</td>
</tr>
<tr>
<td>1976</td>
<td>15311</td>
<td>( L_4 )</td>
</tr>
<tr>
<td>1977</td>
<td>15603</td>
<td>( L_4 )</td>
</tr>
<tr>
<td>1978</td>
<td>15861</td>
<td>( L_5 )</td>
</tr>
<tr>
<td>1979</td>
<td>16807</td>
<td>( L_6 )</td>
</tr>
<tr>
<td>1980</td>
<td>16919</td>
<td>( L_6 )</td>
</tr>
</tbody>
</table>

We use a second-order fuzzy time series. For example, FLRs involves a structure composed of three consecutive fuzzy sets \( A(t-2), A(t-1) \rightarrow A(t) \). Following this concept, two input variables \( A(t-2) \) and \( A(t-1) \) were used as input values for artificial neural networks, and the use of these two input values, it can be assumed fuzzy time series \( A(t) \).

**Step 2. Forecasting results with defuzzification process**

The forecasting values were obtained following defuzzification, and the results were computed using the artificial neural network outputs from Step 1.

4 Application to Enrollment Data

We collected twenty-two years of from the University of Alabama (1971-1992. We first define the universe of discourse \( U \) of the universe by a minimum enrollment number, and the maximum number for enrollment history for use in the initial iteration. FCM-based discretization partitioning was then applied to change the number of clusters between 5 and 20. We then calculated the appropriate FLR by using artificial neural network optimization recommendations. In constructing the fuzzy relationship, the number of neurons in the hidden layer is tuned between 1 and 10, with only one neuron in the input and output layers. Therefore, we implemented a hybrid prediction model with 160 architectures based on 10 different neural network architectures and 16 different clusters.

In this experimental design, the root mean square error (RMSE) value is used as the performance evaluation index by Eq.(4).

\[
RMSE = \sqrt{\frac{\sum_{t=1}^{n} (\text{actual}(t) - \text{forecast}(t))^2}{n}}
\]

\[(4)\]

This research set the linguistic value \( L_i \) of \( A(t-2) \) and \( A(t-1) \) as the input \( \{\text{input-2} \text{ and input-1}\} \) variables for an artificial neural network. The linguistic value \( L_i \) of the \( A(t) \) fuzzy time series can be seen as the forecasting value following computation using the artificial neural network. In Table 3, the input variables for the second observation \([L_1]\) are 1, and then the output variable \( L_2 \) is 2. This FLR is known as the first-order fuzzy time series. Take the third observation as an example, the input variables \([L_1, L_2]\) are 1 and 2, and then the output variable \( L_3 \) is 3. This FLR is known as the second-order fuzzy time series.

**Table 3. FLR construction for fuzzy time series model**

<table>
<thead>
<tr>
<th>Observation no.</th>
<th>( A(t-2) )</th>
<th>( A(t-1) )</th>
<th>( A(t) )</th>
<th>Input-1</th>
<th>Input-2</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>( L_1 )</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>( L_1 )</td>
<td>( L_2 )</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>( L_3 )</td>
<td>( L_2 )</td>
<td>( L_3 )</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>( L_3 )</td>
<td>( L_4 )</td>
<td>( L_4 )</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>( L_4 )</td>
<td>( L_4 )</td>
<td>( L_5 )</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>( L_4 )</td>
<td>( L_5 )</td>
<td>( L_4 )</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>( L_5 )</td>
<td>( L_6 )</td>
<td>( L_7 )</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>
Enrollment forecasting was conducted using several popular models, including those proposed by Song and Chissom (1993b) [19], Song and Chissom (1994) [20], Sullivan and Woodall (1994) [44], Chen (1996) [45], and Cheng et al. (2008) [10]. As shown in Table 4, the proposed models significantly outperformed the traditional time series models.

Table 4. Results comparison

<table>
<thead>
<tr>
<th>Methods</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song and Chissom (1993b)</td>
<td>642.26</td>
</tr>
<tr>
<td>Song and Chissom (1994)</td>
<td>880.73</td>
</tr>
<tr>
<td>Sullivan and Woodall (1994)</td>
<td>621.33</td>
</tr>
<tr>
<td>Chen (1996)</td>
<td>638.36</td>
</tr>
<tr>
<td>Cheng et al. (2008)</td>
<td>478.45</td>
</tr>
<tr>
<td>Proposed method – first order</td>
<td>191.94</td>
</tr>
<tr>
<td>Proposed method – second order</td>
<td>174.03</td>
</tr>
</tbody>
</table>

5 Conclusions

A new high-order fuzzy time series model is proposed using the FCM algorithm to handle discrete partitioning issues, forecasting predicted values based on an artificial neural network optimization method. The proposed model outperforms five conventional fuzzy time series models for forecasting university enrollment. This research presents a high-order fuzzy time series model to solve related time series forecasting problems. However, additional research is considered. First, the model should be applied to additional real-world datasets to provide more performance verification. Second, many curve fitting algorithms can be applied to predict the performance of neural network models. Finally, future researches could deal with discretization partitioning issues by using bio-inspired computing approaches, like particle swarm optimization (PSO), ant colony optimization (ACO), or artificial bee colony (ABC).

Acknowledgments

The authors thank the support of Ministry of Science and Technology (MOST) of the Republic of China (ROC) to this work under Grant No. MOST103-2410-H-025-022-MY2 and MOST103-2410-H-025-017. The authors also gratefully acknowledge the anonymous reviewers for their valuable comments and constructive suggestions.

6 References


