Forecasting Jakarta composite index (IHSG) based on *chen* fuzzy time series and firefly clustering algorithm

R W Ningrum1,*, B Surarso2, Farikhin3 and Y M Safarudin3

1 Magister Program of Mathematics, Department of Mathematics, Faculty of Science and Mathematics, Diponegoro University Semarang, Indonesia
2 Department of Mathematics, Faculty of Science and Mathematics, Diponegoro University Semarang, Indonesia
3 Department of Mechanical Engineering, State Polytechnic of Semarang, Indonesia

*Corresponding author: rianiwidyaningrum@gmail.com

Abstract. This paper proposes the combination of Firefly Algorithm (FA) and *chen* Fuzzy Time Series Forecasting. Most of the existing fuzzy forecasting methods based on fuzzy time series use the static length of intervals. Therefore, we apply an artificial intelligence, i.e., Firefly Algorithm (FA) to set non-stationary length of intervals for each cluster on *chen* Method. The method is evaluated by applying on the Jakarta Composite Index (IHSG) and compare with classical *chen* Fuzzy Time Series Forecasting. Its performance verified through simulation using Matlab.

1. Introduction

In recent years, some methods have been presented based on fuzzy time series predictions in many areas, such as electric load, forecasting stock price, university enrolments, the weather, the number of outpatient visit, etc. When using fuzzy time series for forecasting, it is obvious that the length of intervals in the universe of discourse is important because it can affect the forecasting accuracy rate. However, many implementations of fuzzy forecasting methods used the static length of intervals, i.e., the same length of intervals [1-3]. Many recent researches develop an automatic clustering using arithmetic computation algorithm [4-9]. Chen and Wang [6,7] investigated the effective lengths of intervals in the universe of discourse. On the other hand, the Firefly Algorithm (FA) [10], is an artificial intelligence method which is also commonly used for clustering. We try to improve *chen* Fuzzy Time Series Method using Firefly Clustering Algorithm. Each cluster may have different interval lengths. We evaluate it by applying on the Jakarta Composite Index (IHSG) and compare with classical *chen* Fuzzy Time Series Forecasting. The result of each forecast method will be evaluated by comparing the value of Root Mean Squares Error (RMSE).

2. Methods

2.1 Fuzzy Time Series

Here are some concepts of fuzzy time series. A fuzzy set $A$ of the universe of discourse $U$, $U = \{u_1, u_2, ..., u_n\}$, is defined as formula (1) [1]:

$$ A = \{f_A(u_i)\}, \quad i = 1,2, ..., n $$

(1)
where $f_A$ is the membership function of the fuzzy set $A$, $f_A: U \rightarrow [0,1], f_A(u_i)$ denotes the grade of membership of $u_i$ in the fuzzy set $A$, and $1 \leq i \leq n$.

**Definition 2.1** [4]
Let $Y(t)(t = \ldots, 0,1,2,\ldots)$, a subset of $\mathbb{R}$, be universe of discourse on fuzzy sets $f_i(t)(i = 1,2,\ldots)$. Then, $F(t)$ is called a fuzzy time series on $Y(t)(t = \ldots, 0,1,2,\ldots)$.

From Definition 2.1, we can see that $F(t)$ can be regarded as a linguistic variable and $f_i(t)(i = 1,2,\ldots)$ can be viewed as possible linguistic values of $F(t)$, where $f_i(t)(i = 1,2,\ldots)$ are represented by fuzzy sets. We can also see that $F(t)$ is a function of time $t$.

**Definition 2.2** [5]
If $F(t)$ is caused by $F(t − 1)$ only, i.e., $F(t − 1) → F(t)$, then it can be expressed as formula (2)

$$F(t) = F(t − 1) \cdot R(t, t − 1)$$

(2)

where $R(t, t − 1)$ is the fuzzy relationship between $F(t − 1)$ and $F(t)$, and $F(t) = F(t − 1) \cdot R(t, t − 1)$ is called the first-order forecasting model of $F(t)$. The relation $F(t − 1) → F(t)$ is called a fuzzy logical relationship, where $F(t − 1)$ is the current state and $F(t)$ is the next state.

### 2.2 Firefly Algorithm for Clustering

Basic firefly algorithm is based on firefly phenomena. The less bright firefly will move to brighter firefly. Basic of firefly movement can be calculated by following equation (3) [10]:

$$X_i^{t+1} = X_i^t + \beta e^{−\gamma r_{ij}(X_j − X_i)} + \alpha \epsilon$$

(3)

Where $X_i$ and $X_j$ represent less bright firefly $i$ and brighter firefly $j$ position, $\beta$ is firefly attractiveness factor, and $\gamma$ is light absorption coefficient. Here, $\alpha$ represent random coefficient, and $\epsilon$ represent a random vector. Firefly algorithm is widely used to solve many optimization problems [11]. The standard firefly equations are necessarily used if the firefly position is ranged widely. The simple firefly equation can be written as formula (4):

$$X_i^{t+1} = X_i^t + \beta(X_j − X_i)$$

(4)

### 2.3 Firefly-Fuzzy Forecasting Method

The classical Chen Method used the static length of intervals, i.e., the same length of intervals. We propose an addition algorithm to generate intervals from the historical numerical data. To generate the non-stationary length of intervals, we use Firefly Algorithm for clustering the historical numerical data shown as Figure 1.

**The proposed method for forecasting is present as follows.**

**Step 1.** Let $D_{\text{min}}$ and $D_{\text{max}}$ be the minimum datum and the maximum datum of known historical data. Based on $D_{\text{min}}$ and $D_{\text{max}}$, we define the universe of discourse $U$ as $[D_{\text{min}} − D_1, D_{\text{max}} + D_2]$, where $D_1$ and $D_2$ are two proper positive numbers.

Partition the universe of discourse $U$ into lengthy intervals $u_1, u_2,\ldots, u_n$. Then, generate the interval length automatically using Firefly Algorithm.

**Step 2.** Let $A_1, A_2,\ldots, A_k$ be fuzzy sets which are linguistic values of the linguistic variable. Define fuzzy sets $A_1, A_2,\ldots, A_k$ on the universe of discourse $U$ as formula (5):

$$X_i^{t+1} = X_i^t + \beta(X_j − X_i)$$
\[ A_1 = a_{11}/u_1 + a_{12}/u_2 + \cdots + a_{1m}/u_m \]
\[ A_2 = a_{21}/u_1 + a_{22}/u_2 + \cdots + a_{2m}/u_m \]
\[ \vdots \]
\[ A_k = a_{k1}/u_1 + a_{k2}/u_2 + \cdots + a_{km}/u_m \]

(5)

Where \( a_{ij} \in [0,1], 1 \leq i \leq k, \text{and} \ 1 \leq j \leq m \). The value of \( a_{ij} \) indicates the grade of membership \( u_j \) in the fuzzy sets \( A_i \). Find out the degree of each datum belonging to each \( A_i (i = 1,2,\ldots,m) \). If the maximum membership of one datum is under \( A_k \), then the fuzzification for this datum is treated as \( A_k \). The fuzzy logical relationships are derived based on the fuzzified historical data. Then, the fuzzy logical relationships \( A_j \rightarrow A_k \) means “If datum \( i \) is \( A_j \), then that of the year \( i + 1 \) is \( A_k \)”, where \( A_j \) is called the current state of the data, and \( A_k \) is called the next state of the data (note: the repeated relationships are counted only once).

**Step 3.** Divide the derived fuzzy logical relationships into groups based on the current state of the historical data of fuzzy logical relationships.

**Step 4.** Calculate the forecasted outputs. The calculations are carried out by the following rules:

1. If the fuzzification of datum \( i \) is \( A_j \), and there is only one fuzzy logical relationship (FLR) in the fuzzy logical relationship groups (FLRG) derived in Step 3 in which the current state of the historical data is \( A_j \), which is shown as formula (6):

\[ A_j \rightarrow A_k \]

where \( A_j \) and \( A_k \) are fuzzy sets and the maximum membership value of \( A_k \) occurs at interval \( u_k \), and the midpoint of \( u_k \) is \( m_k \), then the forecasted datum \( i + 1 \) is \( m_k \).

2. If the fuzzification of datum \( i \) is \( A_j \), and there are the following fuzzy logical relationships in the fuzzy logical relationship groups derived in Step 3 in which the current states of the fuzzy logical relationships are \( A_j \), respectively, which is shown as formula (7):

\[ A_j \rightarrow A_{k1} \]
\[ A_j \rightarrow A_{k2} \]
\[ \vdots \]
\[ A_j \rightarrow A_{kp} \]

where \( A_j, A_{k1}, A_{k2}, \ldots, A_{kp} \) are fuzzy sets, and the maximum membership values of \( A_{k1}, A_{k2}, \ldots, A_{kp} \) occurs at intervals \( u_1, u_2, \ldots, u_p \), respectively, and the midpoints of \( u_1, u_2, \ldots, u_p \) is \( m_1, m_2, \ldots, m_p \), respectively, then the forecasted enrollment of datum \( i + 1 \) is \( (m_1 + m_2 + \cdots + m_p)/p \).

3. If the fuzzification of datum \( i \) is \( A_j \), and there does not exist any fuzzy logical relationship groups whose current state of the enrollments is \( A_j \), where the maximum membership value of \( A_j \) occurs at interval \( u_j \), and the midpoint of \( u_j \) is \( m_j \), then the fuzzification of datum \( i + 1 \) is \( m_j \).
This method has a higher accuracy rates than the previous method. In basic firefly algorithm, $\beta$ coefficient represents firefly attractiveness factor that is ranging from 0 to 1. Higher $\beta$ coefficient will guarantee faster computation but less accurate while low $\beta$ coefficient will make the computation slow but more accurate.

In this paper, the $\beta$ coefficient will be updated each iteration to make it faster for convergence, but still accurate. At first, we set $\beta$ coefficient as 0.15. It means that each next iteration firefly will move faster than the last iteration, but still accurate.

3. Result and Discussion
In this section, we apply the proposed method to forecast the known historical data, monthly Jakarta Composite Index (IHSG) from the year 2009 to the year 2017, obtained from idx.co.id.

Let $n$ be the number of clusters, and we set $n = 10$. We compare the result of forecasting between Chen Method and the proposed method. Then, we compare the results of $n = 50$ and $n = 100$.

The method for forecasting is presented as follows.

Step 1: Define the universe of discourse $U$. We can see from the data that $D_{\text{min}} = 1,285.476$ and $D_{\text{max}} = 5,829.708$. Thus, initially, we let $D_1 = 285.476$ and $D_2 = 170.292$. Thus, in this paper the universe of discourse $U = [1000, 6000]$.

Step 2: Partition the universe of discourse $U$ into even lengthy intervals $u_1, u_2, \ldots, u_n$. In this paper, we partition $U = [1000, 6000]$ into ten intervals. Thus, in the other word, the number of clusters for simulation is 10 clusters. Then, we used Firefly Algorithm to clustering the universe of discourse randomly. The algorithm is shown in Figure 1.

Step 3: Let $A_1, A_2, \ldots, A_k$ be the fuzzy sets which are linguistic values of the linguistic values of the linguistic variable Jakarta Composite Index (IHSG) data. In this paper, there are ten linguistic values, as formula (8).

\[
\begin{align*}
A_1 &= 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10} \\
A_2 &= 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10} \\
A_3 &= 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10} \\
&\vdots \\
A_{10} &= 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0.5/u_9 + 1/u_{10}
\end{align*}
\]

The fuzzification of Jakarta Composite Index (IHSG) data is shown in Table 1. Then, Divide the derived fuzzy logical relationships into groups based on the current states of the fuzzy logical relationships, which are shown in Table 2.

| Year | Month | Actual Data | Fuzzification |
|------|-------|-------------|---------------|
| 2009 | Jan   | 1332.667    | $A_1$         |
|      | Feb   | 1285.476    | $A_1$         |
|      | Mar   | 1434.074    | $A_1$         |
|      | \vdots | \vdots | \vdots |
| 2017 | Jan   | 5302.662    | $A_9$         |
|      | June  | 5829.708    | $A_{10}$      |

Step 4: Calculate the forecast outputs. The forecasting results are shown in Table 3. We also try to compare the results for $n = 50$ clusters and $n = 100$ clusters.
Table 2. Fuzzy Logical Relationship Groups (FLRG) of The Data

| Group  | FLRG                      |
|--------|---------------------------|
| 1      | $A_1 \rightarrow A_1$    |
| 2      | $A_2 \rightarrow A_2$    |
| 3      | $A_3 \rightarrow A_3$    |
| 4      | $A_4 \rightarrow A_4$    |
| 5      | $A_5 \rightarrow A_5$    |
| 6      | $A_6 \rightarrow A_6$    |
| 7      | $A_7 \rightarrow A_7$    |
| 8      | $A_8 \rightarrow A_8$    |
| 9      | $A_9 \rightarrow A_9$    |
| 10     | $A_{10} \rightarrow A_{10}$ |

Table 3. Forecasting result

| Year | Month | Actual Data | Forecasting result for $n = 10$ clusters | Forecasting result for $n = 50$ clusters | Forecasting result for $n = 100$ clusters |
|------|-------|-------------|-------------------------------------------|------------------------------------------|------------------------------------------|
|      |       |             | Chen Method                              | Proposed Method                          | Chen Method                              |
|      |       |             | Proposed Method                          | Chen Method                              | Proposed Method                          |
|      |       |             | Proposed Method                          | Chen Method                              | Proposed Method                          |
|      |       |             | Proposed Method                          | Chen Method                              | Proposed Method                          |
| 2009 | Jan   | 1332.667    | *                                        | *                                        | *                                        |
|      |       |             | 1500                                    | 1582.443492                             | 1250                                    |
|      |       |             | 1250                                    | 1369.845689                             | 1275                                    |
|      |       |             | 1289.28669                              |                                          |                                          |
|      | Feb   | 1285.476    | 1500                                    | 1582.443492                             | 1250                                    |
|      |       |             | 1250                                    | 1369.845689                             | 1275                                    |
|      |       |             | 1289.28669                              |                                          |                                          |
|      |       |             |                                          |                                          |                                          |
| 2017 | Jan   | 5302.662    | 5250                                    | 5324.575796                             | 5250                                    |
|      |       |             | 5250                                    | 5366.690661                             | 5400                                    |
|      |       |             | 5383.794854                             |                                          |                                          |
|      | Feb   | 5386.692    | 5250                                    | 5324.575796                             | 5450                                    |
|      |       |             | 5450                                    | 5366.690661                             | 5375                                    |
|      |       |             | 5383.794854                             |                                          |                                          |
|      |       |             |                                          |                                          |                                          |
|      | June  | 5829.708    | 5500                                    | 5785.979614                             | 5850                                    |
|      |       |             | 5850                                    | 5788.8298                              | 5825                                    |
|      |       |             | 5833.838387                             |                                          |                                          |

**IIHS Value**

Figure 1. Forecasting Result for $n = 10$ clusters and Actual Data
From Table 3, we can see that the forecasted results of the proposed method are quite close to the Chen Method. The curves of the actual data and the forecasted results of the proposed method are shown in Figure 2. The blue line is the actual data, the red line is Chen Method forecasting results, and the green line is the forecasted results of the proposed method. It is obvious that the proposed method is more efficient than the Chen Method.

In this paper, the root mean square error (RMSE) is used to compare the forecasted accuracy rates for different methods, shown as formula (9)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\text{Forecasted}_{value_i} - \text{Actual}_{value_i})^2}$$

Table 4. A comparison of RMSE for each method at different number of clusters

| Number of clusters | Chen Method | Proposed Method |
|--------------------|-------------|-----------------|
| 10                 | 228.0103    | 184.7329        |
| 50                 | 144.8957    | 124.0389        |
| 100                | 113.5088    | 101.4061        |

Table 4 shows a comparison of the forecasting root mean square errors (RMSEs) for each method at a different number of clusters. From Table 4, we can see that the proposed method outperforms Chen Method for forecasting the Jakarta Composite Index (IHSG).

4. Conclusion

In this paper, we have presented a Chen Fuzzy Time Series Forecasting with an addition algorithm for clustering, i.e., Firefly Algorithm, to generate the different length of intervals in the universe of discourse. This algorithm allows for optimal interval length for a better forecasting results.

From the experimental results in Table 3 and the comparison of RMSE in Table 4, we can see that the proposed method outperforms classical Chen Method. However, the proposed method requires longer computation if the number of clusters determined is increasing. For future research, we can change the value of membership degree in fuzzy logical relationships (FLRs) to improve the forecasting performance by determining the factors which affect the value of Jakarta Composite Index (IHSG).

References

[1] Bas E, Yoku U, Egrioglu E and Aladag C H 2015 American Journal of Intelligent Systems 5 81
[2] Aripin A, Suryono S and Surarso B 2016 AIP Conference Proceedings 1746 020046
[3] Li Jian and Wang J Q 2017 International Journal of Fuzzy Systems 19 1866
[4] Chen SM, Chu HP and Sheu TW 2012 IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 42 1485
[5] Chen S M and Tanuwijaya K 2011 Expert Systems with Application 38 10594
[6] Chen S M and Phuong B D H 2017 Knowledge-Based Systems 118 204
[7] Wang L, Liu X and Pedrycz W 2013 Expert Systems with Applications 40 5673
[8] Sriparna S and Ranjita D 2017 Neural Computing and Applications 15 1
[9] Wladyslaw H and Jastrzebska A 2017 Neurocomputing 232 3
[10] Long N C, Meesad P and Herwig U 2015 A Expert Systems and Applications 42 8221
[11] Safarudin Y M, Priyadi A, Purnomo M H and Pujiantara M 6th International Conference on Information Technology and Electrical Engineering (ICITEE)