Dynamic Type-2 Fuzzy Time Warping (DT2FTW): A Hybrid Model for Uncertain Time-Series Prediction

Aref Safari¹, Rahil Hosseini¹, and Mahdi Mazinani²

¹Department of Computer Engineering, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran
²Department of Electronic Engineering, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran

Abstract

Prediction of time series is associated with nondeterministic pattern analysis for uncertain conditions. Therefore, it is necessary to develop high-quality prediction methods for real-world applications. Type-2 fuzzy systems can handle high-order uncertainties, such as sequential dependencies associated with time series. Precise and reliable prediction can help to develop reasonable strategies and assist specialists in planning the best policies for modeling events in uncertain time series. In this study, a hybrid model (dynamic type-2 fuzzy time warping [DT2FTW]) was proposed for handling high-order uncertainties in time-series prediction. A type-2 fuzzy intelligent system was developed alongside a dynamic time warping algorithm for predicting the patterns’ similarity in long-time series for time-series prediction. The results demonstrate that the proposed DT2FTW model yields more reliable predictions on global standard benchmarks such as the Mackey-Glass, Dow Jones, and NASDAQ time-series. The results also confirm that the proposed DT2FTW model has lower error rates than its counterpart algorithms in terms of the root mean square error (RMSE), mean absolute error (MAE), and mean percentage error (MPE). In addition, the results confirm the superiority of the proposed model with an average area under the ROC curve (AUC) of 94%, with the 95% confidence interval (92%-95%).

Keywords: Dynamic time warping, Interval type-2 fuzzy system, Time-series prediction

1. Introduction

Modeling of time-series is important because many pattern-analysis problems contain the time component. These problems are typically not addressed, because the time dependence makes time-series-related problems difficult to handle. One of the most challenging issues associated with time-series is their prediction. Prediction problems are often classified into short-term, medium-term, and long-term, and are characterized by different orders of uncertainty. The uncertainty associated with time-series data is implicit, with a nonlinear pattern. On the other hand, unreliable accuracy is a major issue in time-series predictions. Different time-series models, such as fuzzy time-series, have been considered for improving the prediction accuracy.

Many intelligent models have been used for analyzing time-series patterns. Recently, the use of soft computing approaches, such as fuzzy logic, neural networks, simulated annealing, and genetic algorithms, has been reported in the literature on the time-series prediction. These
Table 1. Related fuzzy models for time-series prediction

| Method                  | Limitation                                      | Accuracy (%) |
|-------------------------|-------------------------------------------------|--------------|
| FCM time-series [17]    | No optimal parameter                            | 77.86        |
| Fuzzy logic [18]        | No data reduction                               | 77.18        |
| Fuzzy Markov [19]       | Complexity of model                             | 69.80        |
| Fuzzy time-series [22]  | High-order uncertainty is not modeled           | 77.00        |
| Fuzzy neural [24]       | Weights are fixed                               | 79.24        |
| Fuzzy-PSO [25]          | Limited to mid-range series                     | 80.12        |
| Gustafson-Kessel fuzzy clustering [26] | Insufficient results for long time-series    | 82.93        |
| PSO-fuzzy time-series [27] | Not reliable for long time-series              | 82.45        |
| Fuzzy-NN time-series clustering [18] | Time complexity is high | 84.73        |

approaches have been considered advantageous compared with traditional methods, because they can address nonlinearities and can approximate many types of complex dynamical systems better than linear statistical models. Fuzzy logic models have been widely adopted because of the prevalent uncertainty in the time-series data. Most related studies use the Euclidean distance metric for measuring time-series intervals. The Euclidean distance metric is widely known to be very sensitive to distortions along the time axis, but it has been very popular with many researchers. The ubiquity of the Euclidean distance metric in the face of increasing evidence of its poor accuracy for time-series prediction is almost certainly owing to its ease of implementation and its time and space efficiency. However, in our work, the problem of distortion along the time axis is addressed by dynamic time warping (DTW) alongside a type-2 fuzzy logic approach, for modeling high-order uncertainties in time-series prediction based on the footprint of uncertainty and unequal lengths of time intervals.

1.1 Literature Review

Analysis and prediction of time-series belong to the field of temporal pattern recognition. A time-series corresponds to a stretch of values on the same scale, indexed by a naturally occurring time, as encountered in many applications in the engineering, ecology, economy, medicine, and finance fields [1]. Five concepts are important for time-series: the starting time, pattern similarity, period range, confidence, and the endpoint time. Many techniques [2][8] have been applied for time-series prediction based on the period and data type. The DTW model has been effectively used to automatically deal with time deformations in different time-series ranges with time-dependent data, for pattern recognition and similarity analysis. The DTW approach is currently used in many areas, including online signature matching and handwriting recognition [9], gestures and sign language recognition [10], knowledge discovery, data mining and clustering [11], pattern recognition and data analysis [12], and signal processing [13][14]. In addition, in many real-world applications temporal problems are complex, uncertain, and chaotic [15]. Fuzzy logic is one of the most effective methods for handling uncertainties in dynamic and nonstationary environments [16]. Table 1 lists various hybrid fuzzy models that have been used for predicting time-series [17][22]. Fuzzy systems, especially hybrid fuzzy models, have been very promising for solving complex problems, where a model estimates and predicts the similarity between two time-series in uncertain conditions [18][23][27].

1.2 Highlights

The main objective of this study was to introduce an intelligent dynamic type-2 fuzzy time warping (DT2FTW) model for predicting long-term temporal data in realistic time-series. In this work, the proposed model aims to overcome the drawbacks of the existing methods and offer a more robust, reliable, and accurate model for predicting long time-series using type-2 fuzzy logic. The model is split into two parts: 1) high-order type-2 fuzzy logic-based time-series prediction and 2) DTW. The type-2 time-series prediction consists of several steps, and we applied the operators of Karnik-Mendel (KM) algorithm for defuzzification. Complex min-max composition operators were applied to all predictions. Then, the prediction performance was evaluated using the root mean square error (RMSE), mean absolute error (MAE), and mean percentage error (MPE) metrics, along with the statistical evaluation using the left-tailed T-test.
The remainder of this paper is organized as follows: Section 2 presents the theoretical research background and materials of the proposed model. Section 3 describes the detailed structure of the proposed model. Performance evaluation and experimental results are presented in Section 4, and the paper is concluded in Section 5.

2. Theoretical Background

This section presents a brief overview of DTW, followed by an overview of the concept of the interval type-2 fuzzy set (IT2FS). Finally, relevant mathematical expressions are provided.

2.1 DTW

To obtain a reliable time-series model for real-time applications, DTW and event-DTW (E-DTW) models were applied in [28], for predicting the optimal distance measure between two related time-series. Since the time is aggregated to represent a reasonable estimate, patterns may match a wide variety of actual time-series. Specifically, the pattern-detection task involves searching for a time-series, S, for a pattern, P.

\[ S = S_1, S_2, \ldots, S_i, \ldots, S_n, \quad (1) \]
\[ P = P_1, P_2, \ldots, P_j, \ldots, P_n. \quad (2) \]

Sequences \( S \) and \( P \) can be organized to form an \( m \times n \) grid or plane, where each point in the plane, \((i, j)\), corresponds to an alignment between elements \( S_i \) and \( P_j \). Therefore, \( W \), a warping path, maps the elements of \( S \) and \( P \); this path is called the “distance” between the elements, which must be minimized to yield lower uncertainty.

\[ W = w_1, w_2, \ldots, w_k, \quad (3) \]

where \( W \) is the sequence of grid points, where each \( w_k \) corresponds to a point \((i, j)_k\). To formulate the DTW algorithm in a time-series problem, we need a distance measure between two elements. The distance function \( \delta \) is the degree of the difference in the square of the difference between two data points, as follows:

\[ \delta(i, j) = |S_i - P_j|, \quad (4) \]
\[ \delta(i, j) = (S_i - P_j)^2. \quad (5) \]

The cumulative distance for each path is as follows:

\[ DTW(S, P) = \min_W \left[ \sum_k^t \delta(W_k) \right], \quad (6) \]

where \( \delta \) is the distance measure between two time-series elements. In addition, there are three constraints on an arbitrary warping path \( W = \{w_1, w_2, \ldots, w_k\} \). Element \( w_k \in W \) represents the mapping between \( x_i \) and \( y_j \), that is, \( w_k \in (S, P) \). The three constraints cause element \( W_{k+1} \) only to appear only at the three adjacent positions.

2.2 An Overview of Interval Type-2 Fuzzy Sets

This section provides definitions of type-2 fuzzy sets (T2FSs) and related but essential concepts. The membership function (MF) of a T2FS of a given element is itself a type-1 fuzzy set (T1FS). A T2FS represented as \( \tilde{A} \) is characterized by a type-2 MF \( \mu_{\tilde{A}}(x, u) \) where \( x \in X \) and \( u \in J_x \subseteq [0, 1] \), that is \[ \tilde{A} = \{(x, u), \mu_{\tilde{A}}(x, u) \} | \forall x \in X, \forall u \in J_x \subseteq [0, 1] \}, \quad (7) \]

where \( 0 \leq \mu_{\tilde{A}}(x, u) \leq 1 \), \( X \) is the domain of the fuzzy set, and \( J_x \) is the domain of the secondary MF at \( x \). \( \tilde{A} \) is defined as \[ \tilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{A}}(x, u) = \int_{x \in X} \frac{\int_{u \in J_x} \mu_{\tilde{A}}(x, u)}{n} \leq [0, 1], \quad (8) \]

where \( \int \) represents a union over the admissible \( x \) and \( u \)

\[ \tilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{A}}(x, u) = \int_{x \in X} \frac{\int_{u \in J_x} \mu_{\tilde{A}}(x, u)\mu_{\tilde{A}}(x, u)}{n} \leq [0, 1], \quad (9) \]

where \( x \) is the primary variable, \( J_x \) the interval in \([0, 1]\), is the primary MF of \( x \), \( u \) is the secondary variable, and \( \int_{u \in J_x} \mu_{\tilde{A}}(x, u) \) is the secondary MF at \( x \). Uncertainty about \( \tilde{A} \) is addressed by the union of all of the primary memberships, called the footprint of uncertainty (FOU) of \( \tilde{A} \), that is, \( \text{FOU}(\tilde{A}) \), and given as \[ \text{FOU}(\tilde{A}) = \bigcup_{x \in X} J_x. \quad (10) \]

The FOU for a Gaussian primary MF with an uncertain standard deviation is shown in Figure 3. The FOU is bounded by an upper bound membership function (UMF) \( \mu_{\tilde{A}}^+(x) \), and a lower bound membership function (LMF) \( \mu_{\tilde{A}}^-(x) \), which are T1FSs; consequently, the membership grade of each element of an IT2FS is identified by an interval of \([\mu_{\tilde{A}}^-(x), \mu_{\tilde{A}}^+(x)]\). In the IT2FSs, the UMF and LMF represent the uncertainties of the
The Gaussian MF was used because most real-world variables follow a standard normal distribution with a gradual change of fuzzy memberships from the normalized DT2FTW distances. In this study, the Gaussian MF was applied to compute the membership degree of the input data or measurements of IT2FSs was defined. The Gaussian MF followed the antecedents. The membership degree of the input data or measurements of IT2FSs is to be defined by Eq. (16)

\[ \hat{\mu}_i = \exp \left( -\frac{(x_i - \bar{x})^2}{2\sigma_c^2} \right) \],

where \( \hat{\mu}_i \) is the membership function (or membership grade) of the Gaussian MF, \( x_i \) is the normalized DT2FTW distance of the class query sequence, and \( \bar{x} \) and \( \sigma_c \) stand for the mean and standard deviation, respectively, of the normalized DT2FTW distances. Eq. (15) was used for normalizing the fuzzy memberships

\[ \mu_i = \frac{\hat{\mu}_i}{\sum_i \hat{\mu}_i}, \]

where \( \mu_i \) is the normalized fuzzy membership degree, and \( \hat{\mu}_i \) is the original fuzzy membership.

Step 2 (Inference): To design the inference systems of the DT2FTW model, we applied a TSK type with three inputs and one output; therefore, each input-output variable corresponds to Gaussian MFs and has \( 2^n \) if-then rules. In this study, \( 2^n \) optimal fuzzy rules were applied to the experiments, for improving performance and for minimizing the prediction error rates in different scenarios. As mentioned above, the main goal of DTW is to equate two time-dependent sequences: \( S = (S_1, S_2, ..., S_t, ..., S_N) \).

The mentioned sequences may well be discrete time-series or feature sequences sampled at intermediate points under chaotic and non-stationary conditions. As a predictable value of \( S \) relies on \( S_{t-\ell} \), which is the past value of \( S \), there are \( s \) antecedents in every rule. According to Eqs. (1)-(3), \( X_{(t)} \) is the observed series at time \( t \), \( T_{(t)} \) represents the trend factor, \( S_{(t)} \) represents the seasonal factor, and \( I_{(t)} \) denotes irregular or random factor white noise components with zero mean and constant variance. Hence, an uncertain time-series condition is used for regulating the arrangement between the \( S \) and \( P \) time patterns in an additive form of the time-series. Therefore, this development can be defined by Eq. (16)

\[ X_{(t+1)} = F(x_{(t)}), \]

where \( X_{(t+1)} \in R^d \) is the condition of the system at time step \( t + 1 \), and \( F \) is a nonlinear vector-valued function of the additive form of different types of time-series. On the other hand, \( x_{(t)} \), \( t = 1, ..., N_s \), is a time-ordered set. This measure is distinct in relation to the state \( x_{(t)} \) of the primary scheme, as follows:

\[ x_{(t)} = h_{(t)}(x_{(t)}), \]
where \( h \) is a nonlinear scalar-valued function, and \( \varepsilon \) is a random variable that constitutes the uncertainties and the presence of noise. It is usually assumed that \( \varepsilon(t) \) is strained from a Gaussian MF. The values of \( h(\varepsilon(t)) \) define the time observations in the mid- and long-term. In contrast, for the random factor, the magnitude of epsilon was defined.

**Step 3** (Finding similarity in sliding windows): By implementing the sliding windows, each data center was trained using observations from 0 to \( K - 1 \). The fuzzy inference system was trained with different significance levels to forecast the upper and lower bounds of the prediction intervals of \( S \) and \( P \). In addition, to estimate the optimal values of \( S \) and \( P \), we applied an accumulated cost matrix algorithm as \( C(m,n) \) function, where \( C(m,n) \) is the minimal interval of the distance measure between \( S \) and \( P \). To compare two time-series, we need a local cost measure (LCM) and a local distance measure (LDM), as the functions

\[
C : m \times n \rightarrow R \geq 0.
\]

To compute the LCM for each interval of the sequences \( S \) and \( P \), the cost matrix can be considered as \( C = R_{m \times n} \) distinct, where \( C(m,n) = C(s,p) \).

**Step 4** (Best warping path): To find the best warping path between \( S \) and \( P \), we need to find the path through the grid, as follows:

\[
W = w_1, w_2, ..., w_k,
\]

\[
W_t = (i_t, j_t),
\]

where \( W \) is the warping function at time step \( t \), and the total distance between \( (i_t, j_t) \) has to be minimized. To evaluate the time-normalized distance measures of \( S \) and \( P \), we propose the following equation:

\[
D(S, P) = \left[ \sum_{t=1}^{k} \frac{d(p_t) \cdot W_t}{\sum_{t=1}^{k} W_t} \right],
\]

where \( d(p_t) \) is the distance between \( i_t \) and \( j_t \) and, \( w_t > 0 \) is the weighting factor. To calculate the best alignment path between \( S \) and \( P \), we used

\[
w_0 = \arg\min_W(D(S, P)).
\]

Moreover, at that point, the reduction was calculated using the boundary conditions of the warping window, as follows:

\[
| i_t - j_t | \leq r,
\]

where \( r > 0 \) is the window length. At the end of this step, a weighting coefficient must be chosen. The time-normalized distance between \( S \) and \( P \) is

\[
D(S, P) = \min_W \left[ \frac{\sum_{t=1}^{k} d(p_t) \cdot W_t}{\sum_{t=1}^{k} W_t} \right].
\]

**Step 5** (Measuring uncertainty): Uncertainty in a model affects the model’s confidence and accuracy. This paper presents a model for managing the uncertainty associated with time-series prediction by considering a distribution over data points. This distribution depends on the data points as follows: \( D = \{ X, Y \} \), where \( D \) is the distribution, and \( X, Y \) are the data points of the sliding windows, as follows: \( X = \{ x_1, x_2, ..., x_n \} \), and \( Y = \{ y_1, y_2, ..., y_n \} \). Therefore, the weight distribution after predicting the time-series can be expressed as follows:

\[
p(X, Y),
\]

where \( \omega \) is the weight of the data points. To approximate this distribution, the Bernoulli rate equation must be calculated, as follows:

\[
p(X, Y) \approx \text{Bern}(\omega; \alpha),
\]

where \( \alpha \) is the Bernoulli rate on the weights. Hence, the model uncertainty is the variance of \( T \) Monte Carlo data points, as follows:

\[
(y) = \frac{1}{T} \sum_{t=1}^{T} (y_t - y)^2,
\]

where \( \{ y_t \}_{t=1}^{T} \) is a set of \( T \) outputs of the DT2FTW, as follows:

\[
y = \frac{1}{T} \sum_{t=1}^{T} y_t.
\]

**Step 6** (Defuzzification): Defuzzification is a process by which elements in a fuzzy set are converted into crisps. The defuzzification process was performed to understand the DT2FTW output better. The output processing of the proposed model used the KM algorithm [30]. For this reason, the time steps (data points) must be assigned to a class with the highest membership. During defuzzification, we set up decision rules to guide the process by assigning time steps to crisp sets. A threshold value for MFs was defuzzified and assigned to the class with the highest membership in the last part. Consequently, a group
of time-lagged values in $d_E$ must have delivered adequate data to rebuild the states of a noticeable complex and dynamical system. The centroid method was applied to the defuzzification process. The centroid $C_A$ of the DT2FTW model is the union of the centroids of all its embedded general forms, as follows [32-33]:

$$c(A) = \frac{\sum_{i=1}^{N} x_i \mu_A(x_i)}{\sum_{i=1}^{N} \mu_A(x_i)},$$

$$C_A = \bigcup_{A_c} c(A_c) = [c_1(\hat{A}), c_r(\hat{A})],$$

where $\bigcup$ is the union operation, and

$$c_1(\hat{A}) = \min_{A_c} c(A_c), \quad c_r(\hat{A}) = \max_{A_c} c(A_c),$$

where $c_1(\hat{A})$ and $c_r(\hat{A})$ can be expressed as follows:

$$c_1(\hat{A}) = \frac{\sum_{i=1}^{L} x_i \mu_A(x_i) + \sum_{i=L+1}^{N} x_i \mu_A(x_i)}{\sum_{i=1}^{L} \mu_A(x_i) + \sum_{i=L+1}^{N} \mu_A(x_i)},$$

$$c_r(\hat{A}) = \frac{\sum_{i=1}^{R} x_i \mu_A(x_i) + \sum_{i=R+1}^{N} x_i \mu_A(x_i)}{\sum_{i=1}^{R} \mu_A(x_i) + \sum_{i=R+1}^{N} \mu_A(x_i)},$$

where the switch points $x_L$ and $x_R$, as well as $c_1(\hat{A})$ and $c_r(\hat{A})$, are computed using the KM algorithm [33].

## 4. Performance Evaluation

In this section, the evaluation of the proposed DT2FTW model is presented. First, the metrics for the performance measurements and datasets used in this study are explained. Then, the statistical results, a comparative study, and experimental results are discussed.

### 4.1 Datasets of This Study

This study used well-known existing time-series datasets, including the Mackey-Glass, NASDAQ, and Dow Jones time-series. The selected datasets represent a wide range of uncertainties through the time-series prediction procedure. The proposed DT2FTW model was complemented to acquire an innovative estimator of the predictions, which also permitted us to compute the uncertainties of predictions for noisy Mackey-Glass chaotic time-series. The test of the proposed model is a simulation of time-series data using the following form of the Mackey-Glass nonlinear delay differential equation

$$x(t) = \frac{0.2x(t-\tau)}{1 + x(t-\tau)} - 0.1x(t).$$

In addition, the NASDAQ is the leading United States electronic stock market. It lists around 3,300 companies. We applied 4,250 pairs of data points from the NASDAQ time-series corresponding to the window from 01/12/2018 to 01/12/2020; these data can be downloaded from Yahoo’s live daily data center. The first 3,025 pairs of the data points were used for training, while the remaining 1,225 pairs of the data points were used for validating the DT2FTW model. From the Dow Jones time-series, we used 1,250 pairs of data points, corresponding to the window from 01/03/2019 to 05/01/2020. These data can be downloaded from Yahoo’s live daily data center. The first 1,025 pairs of the data points were used for training, while the remaining 675 pairs of the data points were used for validating the DT2FTW model.

### 4.2 Performance Metrics

To evaluate the prediction error, MAE, RMSE, and MPE metrics were applied to the proposed DT2FTW model, as follows:

$$MAE = \frac{1}{2} \sum_{t=1}^{n} |(S_t - P_t)|,$$

$$RMSE = \sqrt{\frac{1}{2} \sum_{t=1}^{n} (S_t - P_t)^2},$$

$$MPE = 100\% \cdot \frac{1}{n} \sum_{t=1}^{n} \frac{(S_t - P_t)}{S_t},$$

where $S$ denotes the real data points in the time-series, $P$ denotes the DT2FTW predictions for the aggregation model in the final step of the prediction procedure, $n$ is the number of data points, and $t$ is the time variable of the time-series.

### 4.3 Statistical Evaluation

A T-test (left-tailed) was used for estimating the proficiency of the DT2FTW method and its robustness. The null hypothesis was $H_0: \mu_i > \mu_j$ and $H_1: \mu_i < \mu_j$, where $\mu_i$ and $\mu_j$ are the means of for the DT2FTW and the DTW models, for 100 different runs of the cross-validation technique, respectively (Eq. 36). The T-test results in Table 2 reveal the superiority of the proposed DT2FTW model for Mackey-Glass time-series.
Table 2. T-test results for the DT2FTW and DTW

| Fold# | DT2FTW | DTW |
|-------|--------|-----|
| 1     | 0.9089 | 0.6414 |
| 2     | 0.9071 | 0.6149 |
| 3     | 0.9079 | 0.6212 |
| 4     | 0.9109 | 0.6311 |
| 5     | 0.9154 | 0.7063 |
| 6     | 0.9237 | 0.7195 |
| 7     | 0.9341 | 0.7431 |
| 8     | 0.9472 | 0.7621 |
| 9     | 0.9429 | 0.7901 |
| 10    | 0.9503 | 0.7811 |
| Mean  | 0.92484 | 0.70108 |

\[ \mu_i, \mu_j = \frac{1}{10} \sum_{k=1}^{10} AUC_j. \quad (36) \]

4.4 Comparative Study and Experimental Results

The results of learning the NASDAQ, Dow Jones, and Mackey-Glass time-series with added noise are detailed in Table 3, where the average RMSE curves and the acceptance ratios during search are shown in Figures 2 and 3, respectively. The obtained results show that the DT2FTW model achieves state-of-the-art performance for global benchmarks for uncertain time-series. These results show a significant difference between the DT2FTW model and its counterpart DTW model, clustering-DTW model [15], and the fuzzy deep artificial neural network (ANN) model [16]. The results, shown in Table 3, confirm that the proposed DT2FTW model performs better, with lower error rates in terms of the RMSE, MAE, and MPE metrics, for all scenarios.

The results of this study adhere to the uncertainty modeling theory, thus confirming that the proposed model is genuinely more reliable and accurate than its counterpart models in the literature. In addition, the results in Figure 2 show that the proposed DT2FTW model outperforms the clustering and fuzzy deep time-series models, for all scenarios, and that it also outperforms the classical DTW model. Figures 3 and 4 reveal that the proposed model better performs on realistic time series such as the Dow Jones and NASDAQ series, and performs better on Mackey-Glass time-series at different delay rates.

4.5 ROC Curve Analysis

ROC curve analysis was conducted for obtaining a reliable estimate of the DT2FTW model’s performance. The following equations were used for assessing the performance based on the ROC curve analysis of the proposed model. In addition, standard metrics, such as precision, recall, and F-measure, were used for evaluating the proposed DT2FTW model; those metrics were defined as follows:

\[ \text{Precision} = \frac{TP}{(TP + FP)} \times 100\%, \quad (37) \]
\[ \text{Recall} = \frac{TP}{(TP + TN)} \times 100, \quad (38) \]
\[ F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (39) \]

Table 6 shows the comparison of ROC curve analysis results between the proposed DT2FTW model and its counterparts, revealing that the proposed DT2FTW model performs significantly better than its counterpart models. The ROC curve analysis results confirm that the proposed DT2FTW model performs 12% better than the fuzzy clustering model [15], and 4% better than the fuzzy deep ANN [16], in terms of the AUC measure.

4.6 Time Complexity Analysis

The time-series complexity was determined from the number of computation steps required for running an algorithm as a function of the input size. Time was measured in hours, minutes,
Figure 2. Error rate comparison of the DT2FTW model with its counterparts, for global time series: (a) NASDAQ, (b) Dow Jones, and (c) Mackey-Glass.

Figure 3. Error rate of the DT2FTW model, for different noise levels: (a) = 1, (b) = 0.5, and (c) = 0.2, for the Mackey-Glass time series.

Figure 4. Prediction fit results of the DT2FTW model, for different datasets: (a) NASDAQ, (b) Dow Jones, and (c) Mackey-Glass.

Table 4. Comparison of average performance of the DT2FTW model with its counterpart models (unit: %)

| Method       | AUC  | CI       | Recall | Precision | F-measure |
|--------------|------|----------|--------|-----------|-----------|
| Fuzzy-clustering | 82   | 80–83    | 82     | 80        | 81        |
| Fuzzy deep ANN     | 90   | 88–91    | 92     | 91        | 90        |
| DT2FTW            | 94   | 92–95    | 94     | 95        | 93        |
Table 5. Average complexity of the proposed DT2FTW model

| Pseudo-code                  | Average complexity                      |
|------------------------------|------------------------------------------|
| Class type                   | \((6 \times O(1) + O(N \times M) + O(N^2))\) |
| Training time                | \((8 \times O(1) + O(N^2))\)              |
| Calculating the output       | \(O(N)\)                                 |
| Calculating the error        | \(O(1) + O(M \times N)\)                 |
| The output                   | \(O(1) + O(N)\)                          |

Table 6. Time consumption of the DT2FTW model, for the three datasets

| Samples | Mackey-Glass | Dow Jones | NASDAQ   |
|---------|--------------|-----------|----------|
| 36      | 0:00:01      | 0:00:08   | 0:00:14  |
| 48      | 0:00:01      | 0:00:09   | 0:00:23  |
| 60      | 0:00:02      | 0:00:12   | 0:00:29  |
| 112     | 0:00:04      | 0:00:21   | 0:00:37  |
| 224     | 0:00:08      | 0:00:32   | 0:01:05  |
| 448     | 0:00:16      | 0:00:55   | 0:02:23  |
| 1,200   | 0:00:29      | 0:01:22   | 0:03:37  |

and seconds (00:00:00). The results are shown in Tables 5 and 6.

Tables 5 and 6 show the results for all tests on the dataset that has been divided into three different scenarios. The DT2FTW model in the first, second, and third measurements for time complexity had a greater order in several runs with different datasets and different scenarios, namely the Mackey-Glass, Dow Jones, and NASDAQ time-series.

4.7 Discussion and Analysis

The proposed DT2FTW model has more degrees-of-freedom than the type-1 or other related models, because of the FOU parameters in T2FSs, and owing to its potential to model non-uniform time intervals. The proficiency of T2FSs has been proven for high-order uncertainties such as non-stationary events in time-series. In addition, in the proposed model, the problem of distortion of the time axis was addressed by the DTW algorithm alongside a type-2 fuzzy logic approach for modeling high-order uncertainties in time-series prediction using the footprint of uncertainty and unequal length of the time intervals. According to the results of the ROC curve analysis, the proposed DT2FTW model was 12% better than the fuzzy clustering model [15], and 4% better than the fuzzy deep ANN model [16], in terms of the AUC metric. Similarly, the experimental results confirmed that the proposed DT2FTW model has lower error rates than the current methods. The experimental results confirmed the superiority of the DT2FTW model in terms of the RMSE, MAE, and MPE metrics.

5. Conclusion

This study presented the DT2FTW model for predicting uncertain time-series. The DT2FTW model was evaluated by applying it to three global standard datasets. It is a generalized DTW algorithm that takes advantage of both DTW and type-2 fuzzy logic, optimal alignment between time instance paths, and prediction of high-order uncertainties. The experimental results confirm that the proposed DT2FTW model has lower error rates than other state-of-the-art algorithms.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

[1] X. Wang and C. Wang, “Time series data cleaning: a survey,” IEEE Access, vol. 8, pp. 1866-1881, 2020. [https://doi.org/10.1109/ACCESS.2019.2962152]

[2] F. Wang, M. Li, Y. Mei, and W. Li, “Time series data mining: a case study with big data analytics approach,” IEEE Access, vol. 8, pp. 14322-14328, 2020. [https://doi.org/10.1109/ACCESS.2020.2966553]

[3] N. Kanungsukkasem and T. Leelanupab, “Financial latent Dirichlet allocation (FinLDA): feature extraction in text and data mining for financial time series prediction,” IEEE Access, vol. 7, pp. 71645-71664, 2019. [https://doi.org/10.1109/ACCESS.2019.2919993]

[4] D. S. Stoffer and H. Ombao, “Special issue on time series analysis in the biological sciences,” Journal of Time Series Analysis, vol. 33, no. 5, pp. 701-703, 2012. [https://doi.org/10.1111/j.1467-9892.2012.00805.x]

[5] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” Nature Medicine, vol. 25, no. 1, pp. 44-56, 2019. [https://doi.org/10.1038/s41591-018-0300-7]
[6] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: a review and new perspectives,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798-1828, 2013. https://doi.org/10.1109/TPAMI.2013.50

[7] J. H. Bose, V. Flunkert, J. Gasthaus, T. Januschowski, D. Lange, D. Salinas, S. Schelter, M. Seeger, and Y. Wang, “Probabilistic demand forecasting at scale,” Proceedings of the VLDB Endowment, vol. 10, no. 12, pp. 1694-1705, 2017. https://doi.org/10.14778/3137765.3137775

[8] H. Sakoe and S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 26, no. 1, pp. 43-49, 1978. https://doi.org/10.1109/TASSP.1978.1163055

[9] C. C. Tappert, C. Y. Suen, and T. Wakahara, “The state of the art in online handwriting recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 8, pp. 787-808, 1990. https://doi.org/10.1109/34.57669

[10] A. Kuzmanic and V. Zanchi, “Hand shape classification using DTW and LCSS as similarity measures for vision-based gesture recognition system,” in Proceedings of 2007 International Conference on Computer as a Tool (EUROCON), Warsaw, Poland, 2007, pp. 264-269. https://doi.org/10.1109/EURCON.2007.4400350

[11] V. Niennattrakul and C. A. Ratanamahatana, “On clustering multimedia time series data using k-means and dynamic time warping,” in Proceedings of 2007 International Conference on Multimedia and Ubiquitous Engineering (MUE), Seoul, Korea, 2007, pp. 733-738, https://doi.org/10.1109/MUE.2007.165

[12] C. Bahlmann and H. Burkhardt, “The writer independent handwriting recognition system frog on hand and cluster generative statistical dynamic time warping,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 3, pp. 299-310, 2004. https://doi.org/10.1109/TPAMI.2004.1262308

[13] T. Kahveci, A. Singh, and A. Gurel, “Similarity searching for multi-attribute sequences,” in Proceedings 14th International Conference on Scientific and Statistical Database Management, Edinburgh, UK, 2002, pp. 175-184. https://doi.org/10.1109/SSDM.2002.1029718

[14] H. Woo, D. L. Boccelli, J. G. Uber, R. Janke, and Y. Su, “Dynamic time warping for quantitative analysis of tracer study time-series water quality data,” Journal of Water Resources Planning and Management, vol. 145, no. 12, article no. 04019052, 2019. https://doi.org/10.1061/(asce)wr.1943-5452.0001115

[15] Y. Zhang, H. Qu, W. Wang, and J. Zhao, “A novel fuzzy time series forecasting model based on multiple linear regression and time series clustering,” Mathematical Problems in Engineering, vol. 2020, article no. 9546792, 2020. https://doi.org/10.1155/2020/9546792

[16] A. Rosato and M. Panella, “Time series prediction using random weights fuzzy neural networks,” in Proceedings of 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2020, pp. 1-6. https://doi.org/10.1109/FUZZ48607.2020.9177651

[17] E. Egrioglu, C. H. Aladag, and U. Yolcu, “Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c-means and neural networks,” Expert Systems with Applications, vol. 40, no. 3, pp. 854-857, 2013. https://doi.org/10.1016/j.eswa.2012.05.040

[18] S. M. Chen, “Forecasting enrollments based on fuzzy time series,” Fuzzy Sets and Systems, vol. 81, no. 3, pp. 311-319, 1996. https://doi.org/10.1016/0165-0114(95)00220-0

[19] K. Ramadani and D. Devianto, “The forecasting model of Bitcoin price with fuzzy time series Markov chain and Chen logical method,” AIP Conference Proceedings, vol. 2296, no. 1, article no. 020095, 2020. https://doi.org/10.1063/5.0032178

[20] C. H. Cheng, G. W. Cheng, and J. W. Wang, “Multi-attribute fuzzy time series method based on fuzzy clustering,” Expert Systems with Applications, vol. 34, no. 2, pp. 1235-1242, 2008. https://doi.org/10.1016/j.eswa.2006.12.013

[21] C. H. Cheng, T. L. Chen, H. J. Teoh, and C. H. Chiang, “Fuzzy time-series based on adaptive expectation model for TAIEX forecasting,” Expert Systems with Applications, vol. 34, no. 2, pp. 1126-1132, 2008. https://doi.org/10.1016/j.eswa.2006.12.021

[22] R. C. Tsaur, J. C. O. Yang, and H. F. Wang, “Fuzzy relation analysis in fuzzy time series model,” Computers &
Dynamic Type-2 Fuzzy Time Warping (DT2FTW): A Hybrid Model for Uncertain Time-Series Prediction

Aref Safari received his M.Sc. degree in artificial intelligence. He is now working on his Ph.D. thesis to model the uncertainty of non-stationary time-series. His research interests are soft computing, time-series analysis, and pattern analysis.

E-mail: safari.aref@gmail.com

Rahil Hosseini received her Ph.D. degree in computational intelligence from Kingston University. She is a faculty member at Department of Computer Engineering. Her main research interests include pattern recognition, fuzzy modeling and data mining.

E-mail: rahilhosseini@gmail.com

Mahdi Mazinani received his Ph.D. degree in electrical engineering from Kingston University. Currently, he is a faculty member of the Department of Electrical Engineering. His main research interests include probability theory and pattern recognition.

E-mail: mahdi.mazinani@yahoo.com