Parallel-Structure Fuzzy Systems for Time Series Prediction

Min-Soo Kim and Seong-Gon Kong

Abstract

This paper presents a parallel-structure fuzzy system (PSFS) for prediction of time series data. The PSFS consists of a multiple number of fuzzy systems connected in parallel. Each component fuzzy system in the PSFS predicts the same future data independently based on its past time series data with different embedding dimension and time delay. The component fuzzy systems are characterized by multiple-input single-output (MISO) Sugeno-type fuzzy rules modeled by clustering input-output product space data. The optimal embedding dimension for each component fuzzy system is chosen to have superior prediction performance for a given value of time delay. The PSFS determines the final prediction result by averaging the outputs of all the component fuzzy systems excluding the predicted data with the minimum and the maximum values in order to reduce error accumulation effect.

Keywords: Time Series Prediction, Parallel-Structure Fuzzy System, Subtractive Clustering, Chaotic Time Series

1. Introduction

Predicting future behavior of a system based on the knowledge regarding its previous behavior is one of the essential and ultimate objectives of science [13]. There are two basic approaches to prediction: model-based approach and nonparametric method. Model-based approach assumes that sufficient prior information is available with which one can construct an accurate mathematical model for prediction. Once a meaningful model is constructed, one makes prediction of the system's future behavior based on the model. Nonparametric approach, on the other hand, directly attempts to analyze a sequence of observations produced by a system to predict its future behavior. With the model-based approach one cannot expect meaningful prediction result unless the model in use is accurate enough.

Corresponding Author: Seong-Gon Kong is with School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907-1285, U.S.A.
E-mail: skong@purdue.edu

Accurate modeling of complex systems usually involves practical limitations. The model-based approach has a difficulty in that it is not always possible to construct an accurate model. Highly nonlinear systems tend to produce very complicated and irregular signals with a large number of degrees of freedom. Nonparametric methods rely on previously observed data for prediction of future data. Though nonparametric approaches often cannot represent full complexity of real systems, many contemporary prediction theories are developed based on the nonparametric approach because of difficulty in constructing accurate mathematical models.

Signals generated from many practical systems show chaotic behaviors due to inherent nonlinear characteristics of the physical system [3]. Chaotic signals are characterized by the property of sensitivity to small initial perturbations. The fact that chaotic time series are sensitive to initial perturbations makes chaotic time series prediction a difficult task. It is well known that long-term prediction of chaotic signals is hard to achieve since the signals show huge difference for small perturbations. Therefore, many practical prediction algorithms such as linear prediction, neural networks [9], and the adaptive algorithms [1] have been used for short-term prediction or trend analysis of chaotic time series data.

This paper presents a parallel-structure fuzzy system (PSFS) for prediction of chaotic time series data. This approach corresponds to a nonparametric approach since it generates a prediction result based on past observations of the system output. The PSFS consists of a multiple number of component fuzzy systems connected in parallel. Each component fuzzy system in the PSFS predicts future data independently based on its past time series data with different embedding dimension and time delay. The embedding dimension determines the number of inputs of each component fuzzy system. According to the time delay, the component fuzzy system takes inputs at different time intervals. Each component fuzzy system produces separate prediction results for a future data at a specific time index. The PSFS determines the final predicted value as an average of all the outputs of the component fuzzy systems excluding the two extremes, the minimum and the maximum values, in order to reduce error accumulation effect.

Each component fuzzy system contains a small number of multiple-input single-output (MISO) Sugeno-type fuzzy rules [12], which are generated by clustering input-output training data. Fuzzy systems can represent uncertainties involved in the behaviors of complex physical systems easier than conventional prediction algorithms [10,14]. In many cases, fuzzy rules are determined according to experience of human experts or engineering common sense. When a structured knowledge is not available and the information is given in the form of numerical input-output data as in many real-world cases, adaptive clustering algorithms [2] can automatically produce fuzzy rules.

The subtractive clustering algorithm [4,7] produces fuzzy rules for prediction of chaotic time series. Inputs to the fuzzy system in the parallel-structure fuzzy system are determined according to time delay and embedding dimension. Starting at 1, and then increasing by 1 up to 10 for a specific choice of time delay, the optimal embedding dimension is chosen as the value to minimize error performance measures in terms of mean-square error (MSE) and maximum absolute error (MAE) when applied to one-step ahead prediction. This procedure is repeated by changing the time delay from 1 to N, the number of component fuzzy systems. Final prediction result of the PSFS is an average of outputs of all the component fuzzy systems excluding fuzzy system outputs with the maximum and the minimum values. Extensive simulations show that the PSFS with five component fuzzy systems successfully predicts chaotic time series data comparing with the prediction results by a single fuzzy system.

2. Prediction of Time Series

Nonparametric approach of time series prediction is based on the assumptions that future behavior of a time series can be represented by a functional relationship of its previous observations. If $m\tau$ previous input data are given at the kth time step, for a time series data $x(k)$, the τ-step-ahead value $x(k+\tau)$ can be expressed as

$$x(k+\tau) = P[x(k), x(k-\tau), x(k-2\tau), \cdots, x(k-(m-1)\tau)] \tag{1}$$

where $P[\cdot]$ denotes a function that represents input-output relationship of nonparametric time series prediction process. Positive integer m is called embedding dimension and τ is time delay. Future value of a time series can be predicted by an output of a linear or nonlinear function $P[\cdot]$ of $m\tau$ previous input data.

Time series prediction methods can be classified into either one-step-ahead prediction or short-term prediction depending on the fact that predicted values are again used as input values. In one-step-ahead prediction, the predicted value of future data $\hat{x}(k+\tau)$ is expressed by its previous m inputs with time delay τ of the data sequence as in Eq. (2).

$$\hat{x}(k+\tau) = P[x(k), x(k-\tau), x(k-2\tau), \cdots, x(k-(m-1)\tau)] \tag{2}$$

In short-term or long-term prediction of time series data, predicted values of the data are again used as inputs for prediction of future data. Eq. (3) shows short-term or long-term prediction of time series. The predicted value of future data $\hat{x}(k+\tau)$ is expressed according to the data previously predicted $\hat{x}(k), \hat{x}(k-\tau), \cdots, \hat{x}(k-(m-1)\tau)$. Therefore, long-term prediction of chaotic time series based on the data previously predicted is a difficult task since small initial error causes enormous error accumulation effects in future values.

$$\hat{x}(k+\tau) = P[\hat{x}(k), \hat{x}(k-\tau), \hat{x}(k-2\tau), \cdots, \hat{x}(k-(m-1)\tau)] \tag{3}$$

The optimal embedding dimension m is determined for a specific time delay τ. For given τ, error performance measures are calculated for the training data and for the validation data. Validation data is a data set not used in training phase for checking if training result is acceptable. Error performance measures are defined as mean-square error and maximum absolute error calculated from the difference between the one-step-ahead prediction results and real data. For a given time series data, MSE and MAE are computed as one increases possible embedding dimension values for a fixed time delay. This process is repeated for training data and for validation data. The optimal embedding dimension corresponds to the embedding dimension whose error measure is minimized both for training and validation data.

3. Parallel-Structure Fuzzy System

3.1 Configuration of a Parallel-Structure Fuzzy System

A parallel-structure fuzzy system (PSFS) predicts future data according to several prediction mechanisms based on different number and different samples of previous data. The PSFS consists of a multiple number of component fuzzy systems connected in parallel for predicting time series. Figure 1 shows the structure of the parallel-structure fuzzy system. The PSFS contains N component fuzzy systems, $FS_1, FS_2, \cdots, FS_N$ connected in parallel. Each component fuzzy system produces independently predicted values of same future data $\hat{x}(k+r)$ at a time index $k+r$ based on previous data. The PSFS produces the final predicted value according to the N prediction results $\hat{x}_1(k+r),\hat{x}_2(k+r),\cdots,\hat{x}_N(k+r)$ of the N component fuzzy systems.

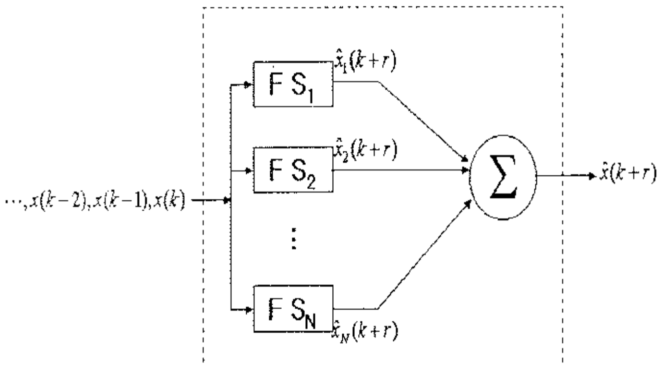


Figure 1. Structure of the parallel-structure fuzzy system(PSFS)

Time series prediction with the PSFS is characterized by the two parameters τ and m. The embedding dimension m defines the number of inputs to each component fuzzy system, and the time delay τ defines the time interval of input data to component fuzzy systems. Time delay τ can have integer values from 1 to N, the number of component fuzzy systems. The optimal embedding dimension for each component fuzzy system must be found for given time delay. Consider a PSFS with N component fuzzy systems. Let τ_i be the time delay and m_i be the embedding dimension for the i th component fuzzy system FS_i, $i=1,\cdots,N$. Then the fuzzy system FS_i is characterized by the parameter pair (τ_i, m_i). In this paper, r is set to 1 and FS_i takes the time delay of $\tau_i = i$, while each component fuzzy system can have arbitrarily different time delay values. Therefore the component fuzzy system FS_i predicts the future value $\hat{x}_i(k+1)$ by means of its m_i previous data as in Eq. (4).

$$\hat{x}_i(k+1)=P[\hat{x}(k+1-i),\hat{x}(k+1-2i),\cdots,\hat{x}(k+1-m_i i)] \tag{4}$$

Figure 2 shows an example of input data used to predict the future data $\hat{x}(k+1)$ using the PSFS. The PSFS consists of three component fuzzy systems FS_1, FS_2, and FS_3. The embedding dimensions for each component fuzzy system are assumed 3, 4, and 3 for the time delay 1, 2, and 3, respectively. So the PSFS is characterized by the 3 parameter pairs (τ_i, m_i) of (1,3), (2,4), and (3,3). This means FS_1 predicts $\hat{x}_1(k+1)$ using the input data $\hat{x}(k),\hat{x}(k-1),\hat{x}(k-2)$, FS_2 generates $\hat{x}_2(k+1)$ using $\hat{x}(k-1)$, $\hat{x}(k-3)$, $\hat{x}(k-5)$, and $\hat{x}(k-7)$. FS_3 outputs $\hat{x}_3(k+1)$ using $\hat{x}(k-2),\hat{x}(k-5),\hat{x}(k-8)$. In the PSFS, final predicted value $\hat{x}(k+1)$ are determined by averaging the 3 prediction results $\hat{x}_1(k+1)$, $\hat{x}_2(k+1)$, and $\hat{x}_3(k+1)$ from the 3 component fuzzy systems. For the PSFS with N component fuzzy systems in general, each fuzzy system produces prediction results $\hat{x}_1(k+1),\hat{x}_2(k+1),\cdots,\hat{x}_N(k+1)$ based on previous data $\hat{x}(k),\hat{x}(k-1),\hat{x}(k-2),\cdots$, and the final predicted data $\hat{x}(k+1)$ becomes an average of all the prediction results by component fuzzy systems.

$$\hat{x}(k+1)=\frac{1}{N-2}\left[\sum_{i=1}^{N}\hat{x}_i(k+1)-\max_i \hat{x}_i(k+1)-\min_i \hat{x}_i(k+1)\right] \tag{5}$$

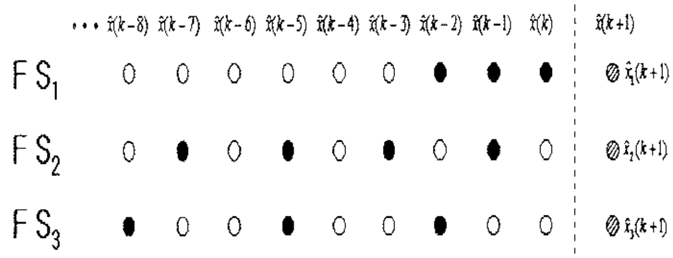


Figure 2. Inputs of each component fuzzy system in the PSFS for prediction of $\hat{x}(k+1)$

In general, consider the PSFS predicting $\hat{x}(k+r)$, the r-step ahead future data at $(k+r)$th time step. The i th component fuzzy system outputs predicted value $\hat{x}_i(k+r)$ based on its previous data $\hat{x}(k+r-i),\hat{x}(k+r-2i),\cdots,\hat{x}(k+r-m_i i)$. The PSFS determines the final predicted value $\hat{x}(k+r)$ as the average of all the component fuzzy system outputs $\hat{x}_1(k+r),\hat{x}_2(k+r),\cdots,\hat{x}_N(k+r)$ excluding the prediction results of the minimum and the maximum values. In short-term or long-term prediction, a small amount of prediction error is accumulated to become a big error after several iterations. The PSFS reduces error accumulation effect by averaging the prediction results of all the component fuzzy systems after removing the extreme values of prediction results.

$$\hat{x}(k+r)=\frac{1}{N-2}\left[\sum_{i=1}^{N}\hat{x}_i(k+r)-\max_i\hat{x}_i(k+r)-\min_i\hat{x}_i(k+r)\right] \quad (6)$$

where

$$\hat{x}_i(k+r)=P[\hat{x}(k+r-i),\hat{x}(k+r-2i),\cdots,\hat{x}(k+r-m_i i)] \quad (7)$$

$i=1,\cdots,N$. Embedding dimension m determines the number of inputs of a component fuzzy system, while inputs to each component fuzzy system are characterized by the time delay. In order to configure the PSFS for time series prediction, the embedding dimension must be determined for specific time delay. For a given time delay, error performance index measures are calculated from the difference between the one-step ahead prediction results and true data. Mean-square error and maximum absolute error are chosen as performance measures. At each choice of time delay, optimal embedding dimension is determined between 1 and 10 as the value at which both the MSE and the MAE of prediction are constant. A constant value of m with smallest MSE and MAE was chosen as the optimal embedding dimension for a fixed τ. Training data is used for modeling the fuzzy system and validation data is for determining the embedding dimension for a fixed time delay. This procedure is repeated until τ increases from 1 to N. The number of component fuzzy systems chosen is five in the simulations below. As the number of component fuzzy systems increases, more accurate prediction results are obtainable, but at extra computational burden. The PSFS produces a final prediction result by averaging the outputs of the component fuzzy systems after removing the maximum and the minimum prediction values.

3.2 Component Fuzzy System

Modeling fuzzy systems involves identification of the structure and the parameters with given training data. In the Sugeno fuzzy model [12], unlike the Mamdani method [10], the consequent part is represented by a linear or nonlinear function of input variables. The Sugeno model can represent nonlinear input-output relationships with a small number of fuzzy rules. Each rule in the Sugeno model corresponds to an input-output relationship of a fuzzy partition. Fuzzy rules in the MISO Sugeno model with n inputs $x_1,\cdots,x_n$ and an output variable y_i of the i th fuzzy rule is of the form:

$$If\ x_1\ is\ A_{i1}\ and\cdots and\ x_n\ is\ A_{in}, Then\ y_i=f_i(x_1,\cdots,x_n) \quad (8)$$

where $i=1,\cdots,M$ and A_{ij} is a linguistic label represented by the membership function $A_{ij}(x_j)$, and $f_i(\cdot)$ denotes a function that relates input to output. M denotes the number of rules in the fuzzy system. The Sugeno fuzzy model can easily generates fuzzy rules from numerical input-output data obtained from an actual process. The function $A_{ij}(x_j)$ defines a membership function assigned to the input variable x_j in the i th fuzzy rule. In this paper, Gaussian membership functions and a linear function are used for simplicity.

$$A_{ij}(x_j)=\exp\left(-\frac{1}{2}\left(\frac{x_j-c_{ij}}{w_{ij}}\right)^2\right) \quad (9)$$

$$f_i(x_1,\cdots,x_n)=a_{0i}+a_{1i}x_1+\cdots+a_{ni}x_n \quad (10)$$

where the parameters c_{ij} and w_{ij} define center and width of the Gaussian membership function $A_{ij}(x_j)$. The coefficients $a_{0i},a_{1i},\cdots,a_{ni}$ are to be determined from input-output training data. In the simplified reasoning method [14], the output y of the fuzzy system with M rules is represented as

$$y=\frac{\sum_{i=1}^{M}\mu_i f_i(x_1,\cdots,x_n)}{\sum_{i=1}^{M}\mu_i} \quad (11)$$

where μ_i is a degree of relevance. In the product implication method, the degree of relevance is defined as

$$\mu_i=\prod_{j=1}^{n}A_{ij}(x_j) \quad (12)$$

$$=\exp\left(-\frac{1}{2}\sum_{j=1}^{n}\left(\frac{x_j-c_{ij}}{w_{ij}}\right)^2\right) \quad (13)$$

In the Sugeno fuzzy model, the time-consuming rule extraction process from experience of human experts or engineering common sense reduces to a simple parameter optimization process of coefficients $a_{0i},a_{1i},\cdots,a_{ni}$ for a given input-output data set since the consequent part is represented by a linear function of input variables. Using linear function in the consequent part, a group of simple fuzzy rules can successfully approximate nonlinear characteristics of practical complex systems [11].

Construction of the parallel-structure fuzzy system is basically off-line. The parameters of the PSFS must be determined by the training data before time series prediction operation. If a sufficient amount of past data is given, it is easy to construct and train a PSFS for prediction of future time series data. In order to construct the PSFS on line, the parameters of each component fuzzy system must be computed each time, which will require a large amount of computational burden. Also on-line training of the PSFS based on chaotic time series data will result in an inaccurate predictor because of error accumulation.

3.3 Modeling of Component Fuzzy System based on Clustering

The parameters of the component fuzzy systems are characterized using clustering algorithms. The subtractive clustering algorithm [7] finds cluster centers $x_i^* = (x_{1i}^*, \cdots, x_{ni}^*)$ of data in input-output product space by computing the potential values at each data point. The potential value is inversely proportional to distance between data points, which means densely populated data produces large potential values and therefore more cluster centers.

In the Subtractive clustering algorithm, the first cluster center corresponds to the data with the largest potential value. After removing the effect of the cluster center just found, the next cluster center becomes the data with the largest potential value, and so on. This procedure is repeated until the potential value becomes smaller than a predetermined threshold. There are n -dimensional input vectors $x_1, x_2, \cdots, x_m$ and 1-dimensional outputs $y_1, y_2, \cdots, y_m$ forming $(n+1)$ -dimensional space of input-output data. For data $X_1, X_2, \cdots, X_m$ in $(n+1)$ -dimensional input-output space, the subtractive clustering algorithm produces cluster centers as in the following procedure:

Step 1: Normalize given data into the interval [0,1].

Step 2: Compute the potential values at each data point. The potential value P_i of the data X_i is computed as

$$P_i = \sum_{j=1}^{N} \exp(-\alpha \|X_i - X_j\|^2), \quad i = 1,2,\cdots,m \tag{14}$$

where a positive constant $\alpha = 4/r_a^2$ determines the data interval which affects the potential values. Data outside the circle with radius over positive constant $r_a < 1$ do not substantially affect potential values.

Step 3: Determine the data with the largest potential value P_1^* as the first cluster center X_1^*.

Step 4: Compute the potential value P_i' after eliminating the influence of the first cluster center.

$$P_i' = P_i - P_1^* \sum_{j=1}^{N} \exp(-\beta \|X_i - X_1^*\|^2) \tag{15}$$

where positive constant $\beta = 4/r_b^2$ prevents the second cluster center from locating close to the first cluster center. If the effect of potential of the first cluster center is not eliminated, second cluster center tends to appear close to the first cluster center, since there are many data concentrated in the first cluster center. Taking $r_b > r_a$ makes the next cluster center not appear near the present cluster center.

Step 5: Determine the data point of the largest potential value P_2^* as the second cluster center X_2^*. In general, compute potential values P_i' after removing the effect of the k th cluster center X_k^*, and choose the data of the largest potential value as the cluster center X_{k+1}^*

$$P_i' = P_i - P_k^* \exp(-\beta \|X_i - X_k^*\|^2) \tag{16}$$

Step 6: Check if we accept the computed cluster center. If $P_k^* / P_1^* \geq \bar{\varepsilon}$, or $P_k^* / P_1^* > \underline{\varepsilon}$ and $\frac{d_{\min}}{r_a} + \frac{P_k^*}{P_1^*} \geq 1$, then accept the cluster center and repeat step 5. Here $d_{\min}$ denotes the shortest distance to the cluster centers $X_1^*, X_2^*, \cdots, X_k^*$ determined so far. If $P_k^* / P_1^* > \underline{\varepsilon}$ and $\frac{d_{\min}}{r_a} + \frac{P_k^*}{P_1^*} < 1$, then set the X_k^* to 0 and select the data of the next largest potential. If $\frac{d_{\min}}{r_a} + \frac{P_k^*}{P_1^*} \geq 1$ for the data, choose this data as the new cluster center and repeat step 5. If $P_k^* / P_1^* \leq \underline{\varepsilon}$, terminate the iteration.

When determining cluster centers, upper limit $\bar{\varepsilon}$ and lower limit $\underline{\varepsilon}$ allows the data of lower potential and of larger distance $d_{\min}$ between cluster centers to be cluster centers. Step 6 is the determining process of the calculated cluster center according to $d_{\min}$, the smallest distance to the cluster centers $X_1^*, X_2^*, \cdots$ calculated so far. When determining the cluster centers, data with low potential value can be chosen as a cluster center if $d_{\min}$ is big enough due to upper limit $\bar{\varepsilon}$ and lower limit $\underline{\varepsilon}$.

Fuzzy system modeling process using the cluster centers $X_1^*, X_2^*, \cdots, X_M^*$ in input-output space is as follows. The input part of the cluster centers corresponds to antecedent fuzzy sets. In $(n+1)$ -dimensional cluster center X_i^*, the first n values are n -dimensional input space $x_i^* = (x_{i1}^*, \cdots, x_{in}^*)$. Each component determines the center of membership functions for each antecedent fuzzy sets. The cluster centers become the center of the membership functions $c_{ij} = x_{ij}^*$. The width of the membership function w_{ij} is decided as

$$w_{ij} = r_a \left\| \max_i(x_i^*) - \min_i(x_i^*) \right\| / \sqrt{M} \quad (17)$$

where M denotes the number of cluster centers, $\left\| \max_i(x_i^*) - \min_i(x_i^*) \right\|$ denotes the difference between the maximum and the minimum distances between cluster centers. The number of cluster centers corresponds to the number of fuzzy rules. The next process is to compute optimal consequent parameters $a_{0i}, a_{1i}, \cdots, a_{ni}$ in order to produce output y_i of the i th rule in the Sugeno fuzzy model. The number of centers equals the number of fuzzy rules. The output of the fuzzy system is defined as a linear function of

$$y_i = a_{0i} + a_{1i}x_1 + a_{2i}x_2 + \cdots + a_{ni}x_n \quad (18)$$

$$= [a_{1i}, a_{2i}, \cdots, a_{ni}]x + a_{0i} \quad (19)$$

$$= a_i^T x + a_{0i} \quad (20)$$

Compute parameters a_i through linear least-squares estimation, the final output y of the Sugeno fuzzy model is given as

$$y = \frac{\sum_{i=1}^{M} \mu_i (a_i^T x + a_{0i})}{\sum_{i=1}^{M} \mu_i} \quad (21)$$

This is the final output of the fuzzy system.

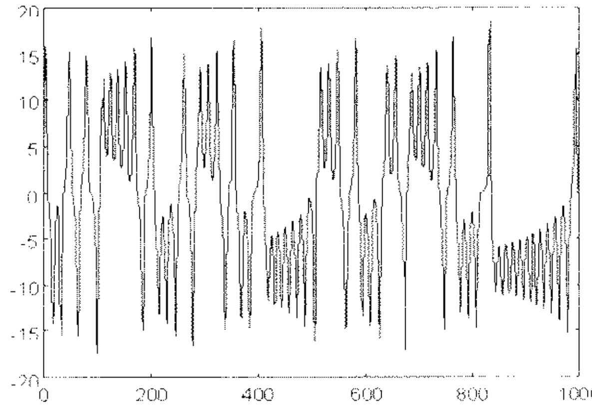
4. Simulation

4.1 Chaotic Time Series Data

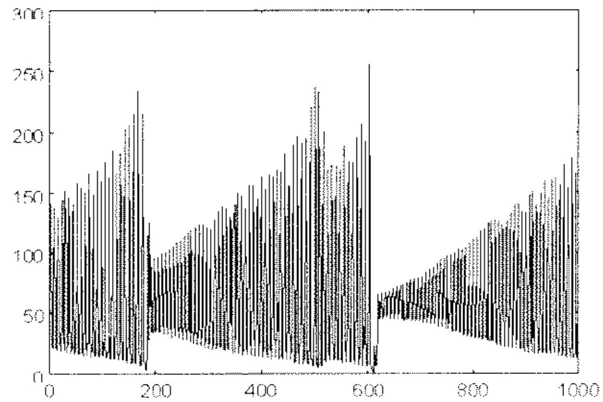
Time series prediction with the parallel-structure fuzzy system was demonstrated using the two chaotic time series data. Lorenz time series data [13] and a signal generated from a laser [6]. The Lorenz time series data used in this paper is x-component solution of the Lorenz equation given as

$$\dot{x} = a(y - x)$$
$$\dot{y} = bx - y - xz \quad (22)$$
$$\dot{z} = -cz + xy$$

where the parameters are set at the values $a = 10, b = 20, c = 20$, and the sampling interval is 0.05. Figure 3 shows x-component Lorenz time series data and the laser time series data used in the prediction with the PSFS. Each data set contains 1000 samples.



(a) Lorenz time series (x-component).



(b) Laser time series

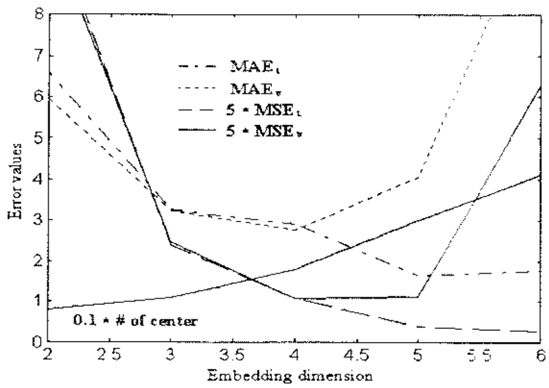
Figure 3. The Lorenz and the laser time series data.

The subtractive clustering algorithm with the parameters $r_a = 0.3, r_b = 0.75, \bar{\varepsilon} = 0.3$, and $\underline{\varepsilon} = 0.1$ generates the cluster centers. The PSFS with 5 component fuzzy systems ($N = 5$) is applied to time series prediction with 1,000 data and with 10,000 data. The prediction performance is compared with a single fuzzy system with the optimal embedding dimension and the time delay. For the prediction with 1,000 data, 500 data were used to train the PSFS, 300 data for validation, and 200 data for test. Second, 10,000 data were used in the training procedure, 200 data were used for test.

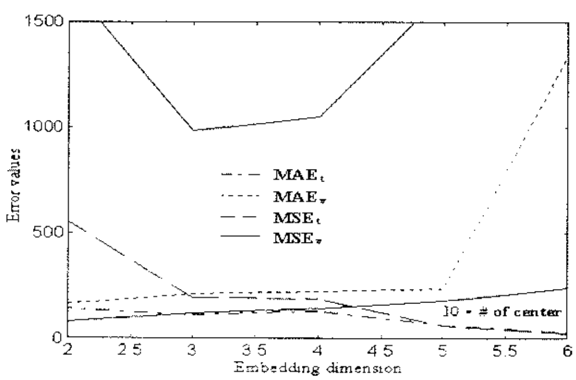
4.2 Prediction with a Single Fuzzy System

In order to configure the PSFS for time series prediction, the embedding dimension m must be determined for a specific time delay τ . For given τ , error performance measures are calculated from the difference between the one-step ahead prediction results trained with training data and validation data. The optimal embedding dimension corresponds to the error measure that is minimized as embedding dimension is increased by 1, starting at 1.

Figure 4 shows how to find the optimal embedding dimension m for given time delay τ. In Figure 4(a), error performance measures are MSE_t, MAE_t, MSE_v, and MAE_v. for training and validation data in Lorenz time series when $\tau = 4$. The optimal value of m becomes 4. This value is very close to the τ calculated by the mutual information [5] or by the false nearest neighbor (FNN)[8] algorithm. For the Lorenz time series, the performance measures reach their minimum at $m = 4$ for $\tau = 4$. For the laser time series, the optimal $m = 3$ for $\tau = 3$.



(a) Lorenz

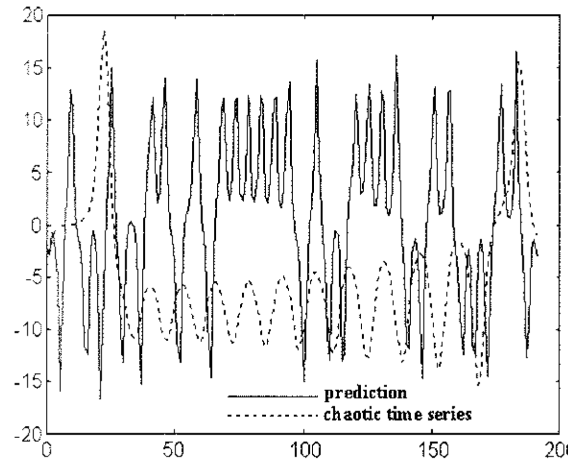


(b) Laser

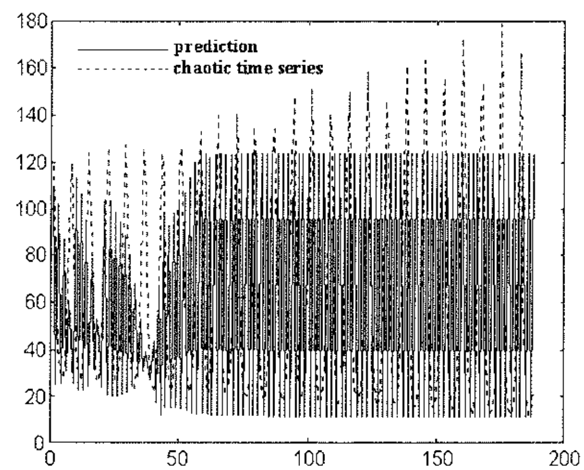
Figure 4. Determination of the optimal embedding dimension

Lorenz time series data is characterized by the five (τ, m) pairs of (1,3), (2,3), (3,3), (4,4), and (5,5). For the laser time series data, the pairs were (1,2), (2,4), (3,3), (3,3), and (5,3). The embedding dimension and the time delay pair (4,4) in the Lorenz and the pair (3,3) in the laser time series data correspond to the time delay determined by the mutual information by Fraser [5] and the minimum embedding dimension computed by the false nearest neighbor (FNN) algorithm.

Figure 5 shows the prediction results by a single fuzzy system using the optimal parameter pair of embedding dimension m and time delay τ. The optimal value of m at a fixed τ corresponds to an integer for which the performance measures MSE and MAE is the smallest value of one step ahead prediction for both training and validation data. Training data are used in constructing the fuzzy system based on the clustering algorithm. Validation data, which is not used to construct the fuzzy system, determines the optimal m according to τ when applied with the one-step-ahead prediction method. Figure 5(a) shows the prediction result of a single fuzzy system for the Lorenz time series with $m = 4$ and $\tau = 4$. Figure 5(b) shows the prediction result for the laser time series with $m = 3$ for $\tau = 3$.



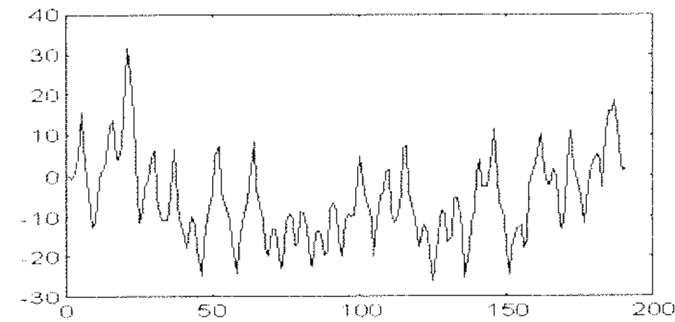
(a) Lorenz



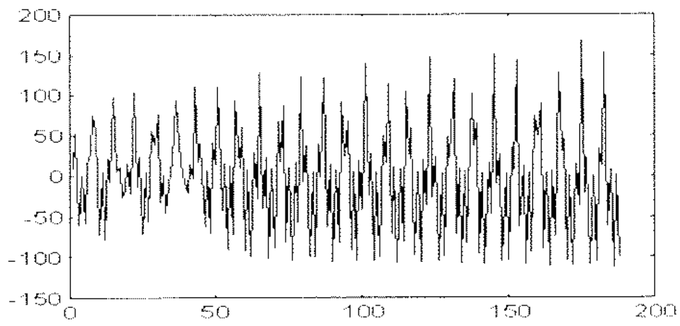
(b) Laser

Figure 5. Prediction result of a single fuzzy system

Figure 6 shows the prediction error defined as the difference between actual values and prediction values. As a result, the single fuzzy system did not produce satisfactory prediction results and is not suitable for time series prediction. The simulation results show that a single prediction system with one τ and m is not suitable for prediction of chaotic time series and it does not produce enough prediction accuracy.



(a) Lorenz



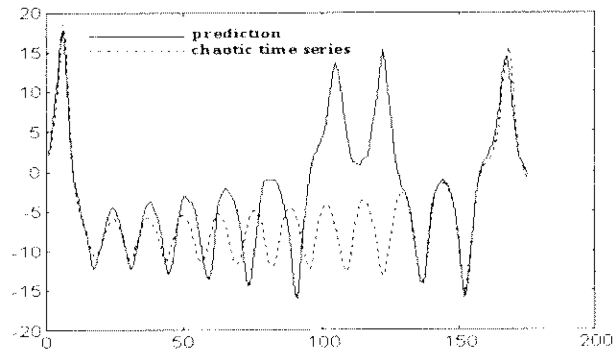
(b) Laser

Figure 6. Prediction error of a single fuzzy system

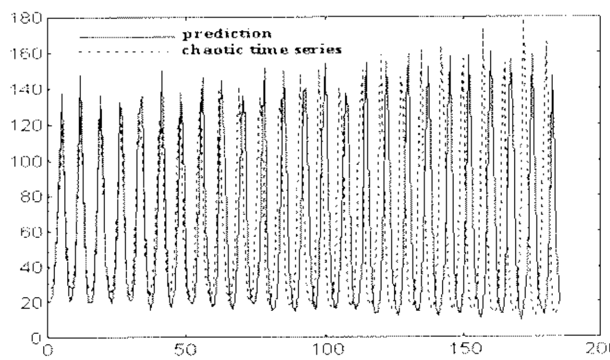
4.3 Prediction with Parallel-Structure Fuzzy Systems

The PSFS is used for prediction of chaotic time series. The PSFS contains five component fuzzy systems ($N = 5$). Each component fuzzy system is characterized by the optimal embedding dimension for corresponding time delay. Five prediction results produced by the component fuzzy systems are averaged excluding the minimum and the maximum prediction values at each step. The maximum and the minimum prediction results are not considered in the final prediction result since it is likely to increase prediction errors.

A 1,000 x-component Lorenz time series data are generated in order to verify the prediction performance of the PSFS. The first 500 data are used for modeling the PSFS with clustering algorithm and for deciding the optimal embedding dimension. The next 300 data are used for validation only to decide the optimal value of m, and the last 200 samples are used to test prediction using the PSFS. The PSFS with $N = 5$ predicts the x-component data of the Lorenz time series using the optimal values of (τ, m) pairs of (1,3), (2,3), (3,3), (4,4), and (5,5). Figure 7 shows the prediction result by the PSFS trained with 500 samples. Figure 8 shows prediction error with the PSFS for the two time series data. In Figure 8, lack of training data causes small prediction errors after 130 iterations in the Lorenz data. This means the 500 training data are not sufficient enough to configure the PSFS for prediction of the chaotic time series data.

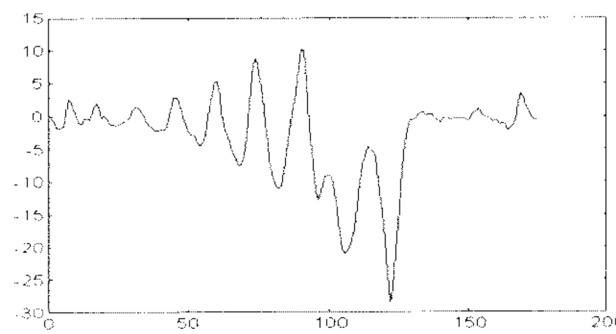


(a) Lorenz

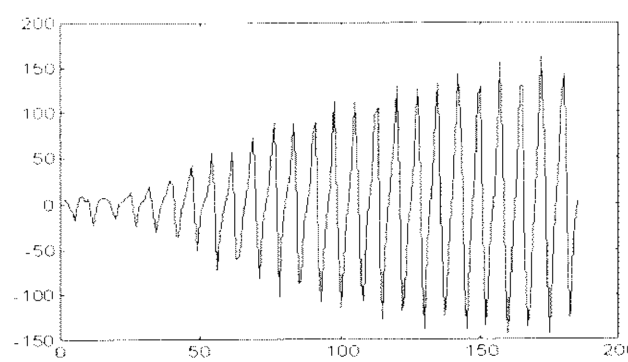


(b) Laser

Figure 7. Prediction result of the PSFS with 500 training samples



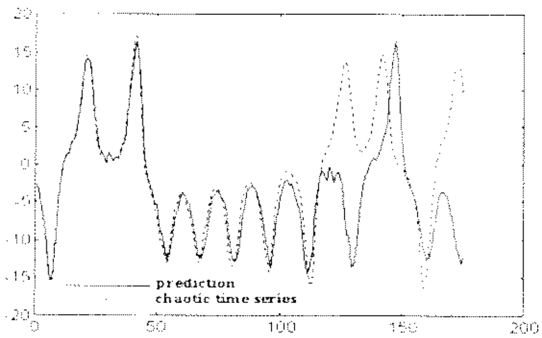
(a) Lorenz



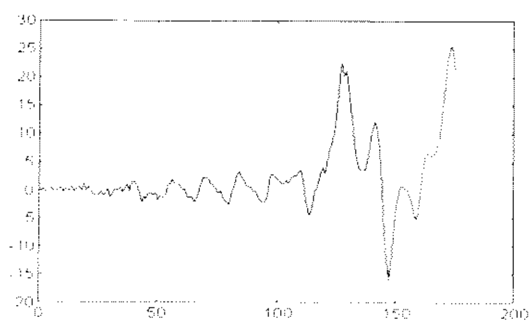
(b) Laser

Figure 8. Prediction error of the PSFS with 500 training samples

Next the number of training and validation data is increased to 10,000. Figure 9 shows the prediction result of the PSFS trained with 10,000 training data. The trained PSFS is applied to predict future data that were not used in the training procedure. The PSFS successfully predicts future data without large error up to 120 samples. This fact shows that the PSFS trained with a sufficient number of data can predict chaotic time series successfully.



(a) Prediction result



(b) Prediction error

Figure 9. Prediction result of the PSFS with 10,000 training samples

5. Conclusion

This paper presents a parallel-structure fuzzy system (PSFS) for predicting chaotic time series. The PSFS corresponds to a nonparametric approach of time series prediction. The PSFS consists of a multiple number of component fuzzy systems connected in parallel. Each component fuzzy system is characterized by multiple-input single-output Sugeno-type fuzzy rules, which are useful for extracting information from numerical input-output training data. The component fuzzy systems for time series prediction are modeled by clustering input-output data. Each component fuzzy system predicts the future value at the same time index with different values of embedding dimension and time delay. The PSFS determines the final prediction value by averaging the results of each fuzzy system excluding the minimum and the maximum values out of N outputs in order to reduce error accumulation effect. Model-based approach was not considered since it has a difficulty in that it is not always possible to construct an accurate model. Performance comparison with the model-based approach will depend on modeling accuracy.

Computer simulations show that the PSFS trained with a training and validation data successfully predicts the x-component Lorenz data and the laser time series data. The PSFS trained with sufficient number of data produces precise prediction results up to 120 time steps for the Lorenz time series data. The embedding dimension determines the number of inputs of a component fuzzy system, and the inputs to each component fuzzy system are characterized by the time delay. The number of component fuzzy systems chosen is five in this simulation. As the number of component fuzzy systems increases, more accurate prediction results are obtainable, but at extra computational burden. At each choice of time delay, optimal embedding dimension is determined by the value at which both the mean-square prediction error and the maximum absolute prediction error are constant. Input to each component fuzzy system are determined by the embedding dimension. The PSFS produces the final prediction result by averaging the outputs of the component fuzzy systems after removing the maximum and the minimum prediction values.

References

[1] D. S. Broomhead and D. Lowe, "Multi-variable Functional Interpolation and Adaptive Networks," *Complex Systems*, Vol. 2, No.3, pp.321-355, 1988.

[2] J. Bezdek, R. Hathaway, M. Sabin, and W. Tucker, "Convergence Theory for Fuzzy c-means: Counter Examples and Repairs," *The Analysis of Fuzzy Information*, J. Bezdek (ed.), Boca Raton: CRC Press, Vol. 3, Ch. 8, 1987.

[3] M. Casdagal, "Nonlinear Prediction of Chaotic Time Series," *Physica D*, pp.335-356, 1989.

[4] S. Chiu, "Fuzzy Model Identification Based on Cluster Estimation," *Journal of Intelligent and Fuzzy Systems*, Vol. 2, No. 3, Sept. 1994.

[5] A. M. Fraser and H. L. Swinney, " Independent Coordinates for Strange Attractors from Mutual Information," *Physical Review A*, Vol. 33, No. 2, pp.1134-1140, 1986.

[6] U. Huebner, N. B. Abraham, and C. O. Weiss, "Dimensions and Entropies of Chaotic Intensity Pulsations in a Single-mode Far-infrared NH3 Laser," *Physical Review A*, Vol. 40, No. 11, pp.6354-6365, 1989.

[7] J.-S. R. Jang, and C.-T. Sun, "Neuro-Fuzzy

Modeling and Control," *Proceedings of the IEEE*, Vol. 83, No. 3, pp.378-406, March 1995.
[8] M. B. Kennel et. al, "Determining Embedding Dimension for Phase Space Reconstruction using Geometrical Construction," *Physical Review A*, Vol. 45, No. 6, pp.3403-3411, 1992.
[9] D. Lowe and A. R. Webb, "Time Series Prediction by Adaptive Networks: A Dynamical Systems Perspective," *IEE Proceedings F*, Vol. 138, No. 1, pp.17-24, 1991.
[10] E. H. Mamdani, and S. Assilian, "An Experiment in Linguistic Synthesis with a Fuzzy Logic Conroller," *Int. J. of Man Machine Studies*, Vol. 7, No. 1, pp.1-13, 1975.
[11] S. Shao, "Fuzzy Self-organizing Control and its Applications for Dynamical Systems," *Fuzzy Sets and Systems*, Vol. 26, pp.151-164, 1988.
[12] M. Sugeno, *Industrial Applications of Fuzzy Control*, Elsevier Science Pub., 1985.
[13] A. S. Weigend and N. A. Gershenfeld, eds., *Time Series Prediction: Forecasting the Future and Understanding the Past*, Massachusetts: Addison-Wesley Pub., pp.175-193, 1994.
[14] R. R. Yager and D. P. Filev, *Essentials of Fuzzy Modeling and Control*, New York: John Wiley & Sons, pp.246-264, 1994.

Min-Soo Kim received the B.S. and the M.S. degrees in electrical engineering from the Soongsil University, Seoul, Korea, in 1995 and 1997, respectively. He is currently working toward the Ph.D. degree at Soongsil University, Seoul, Korea. His research interests include time series prediction, cooperative multi-agent learning, fuzzy systems, and artificial neural networks.

Seong-Gon Kong received the B.S. and the M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1982 and 1987. In 1991 he received the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, California.

Dr. Kong is an Associate Professorof Electrical Engineering at Soongsil University, Seoul, Korea, and was Department Chair from August 1998 to July 2000. From August 2000 he is with School of Electrical and Computer Engineering at Purdue University, West Lafayette, Indiana as a visiting scholar. His research interests include artificial neural networks, fuzzy systems, pattern recognition, and evolutionary robotics.