Fuzzy Time Series Prediction with Data Preprocessing and Error Compensation Based on Correlation Analysis

Young-Keun. Bang, Chul-Heui. Lee

Dept. of Electrical and Electronics Engineering, Kangwon National University
192-1 Hyojadong, Chunchon, Kangwondo, Korea
b2y2c1@hanmail.net chlee@kangwon.ac.kr

Abstract

In general, it is difficult to predict non-stationary or chaotic time series since there exists drift and/or non-linearity as well as uncertainty in them. To overcome this situation, we propose an effective prediction method which adopts data preprocessing and multiple model TS fuzzy predictors combined with model selection mechanism. The proposed method uses the differences of time series as predictor input instead of the original ones because the difference data can stabilize the statistical characteristics of those time series and reveals better their implicit properties. In data preprocessing procedure, the candidates of optimal difference interval are determined based on the correlation analysis, and corresponding difference data sets are generated for them. And then, TS fuzzy predictors are constructed for multiple model bank, where k-means clustering algorithm is used for fuzzy partition of input space, and the least squares method is applied to parameter identification of fuzzy rules. Among the predictors in the model bank, the best one which minimizes the performance index is selected, and it works on hereafter for prediction. Finally, the error compensation procedure based on correlation analysis is added to improve the prediction accuracy. Computer simulation on some typical examples is performed to verify the effectiveness of the proposed method.

1. Introduction

Time series is the data which is observed at regular intervals, and it is generated in various fields concerned with our daily life such as weather, stock market, and electric power, etc. The essential goal of time series analysis is to predict future behavior of a system based on the knowledge about its previous behavior. To predict time series, first of all it is necessary to analyze them. Time series analysis is to find out the various probabilistic and statistical characteristics in them. Also, this is to make models for a prediction system using the tendency, circulation, seasonal change and probabilistic change in them. Generally, Linear regression models (such as AR, MR and ARMA) and Box-Jenkins' model (like ARIMA) have been most widely used for time series analysis. [1][2] Recently, many researchers have been considering the soft computing techniques such as fuzzy theory, neural networks and genetic algorithm. [3-5] especially, fuzzy theory is one of most useful technique to handle the approximate and vague characteristics. However, although traditional fuzzy methods of using original time series have resulted in good performance on stationary time series prediction, they have brought about wrong performance on non-stationary time series prediction because these methods could not handle the statistical properties (such as the mean and dispersion) of non-stationary time series properly which continuously change with time. Hence, to overcome this problem the methods that use differences of time series to predictor input had been proposed. [6-7] these methods were able to not only stabilize the statistical properties but also improve the prediction performances of a system. However, the above paper [6] has indicated restrictive system performance because it had used only first-order difference of time series. In the other paper [7], we have improved the problem of the paper [6] using various differences but it still needed to research about how to detect the candidates of difference which reveal the patterns and rules in non-stationary time series appropriately. Thus, in this paper, we propose more effective methods for predicting them. Firstly, we propose data preprocessing more useful than previous paper. In this procedure, we used correlation analysis to find the similarities between differences and original

978-0-7695-3407-7/08 $25.00 © 2008 IEEE
DOI 10.1109/ICCIT.2008.302

time series and used subtraction between neighbor coefficients to detect the number of the candidate. Through proposed procedure, we detected the candidates of optimal difference interval that reveal the properties of time series properly, and generated difference data sets corresponding to them for the prediction system. And then we constructed multiple model bank using TS fuzzy predictors which use the above differences. For fuzzy partition of input space, and parameter identification of fuzzy rules, we used k-means clustering algorithm [8] and applied the least squares method. Among the predictors in the model bank, the best one which minimizes the performance index was selected, and we allowed it to perform prediction for the rest. Finally, we added the error compensation procedure based on correlation analysis to improve the prediction accuracy. Also we verified the effectiveness of our method through computer simulation on some typical examples.

2. Concept of proposed fuzzy time series prediction

TS fuzzy model is widely used in prediction field, and its rule is defined such as formula (1). In formula (1), TS fuzzy model can get advantage of classical linear-regression model from linear formula and efficiently accept uncertainty and non-linearity by fuzzy sets. Therefore, we constructed fuzzy prediction system using the TS fuzzy model.

$$\begin{aligned} &R: If\ x_1\ is\ A_1\ and\ \ x_2\ is\ A_2\ and \cdots and\ \ x_n\ is\ A_n \\ &\quad Then\ \ y = a_1\,x_1 + a_1\,x_1 + \cdots + a_n x_n \end{aligned} \tag{1}$$

In formula (1), it is necessary to divide input space about given information and data, and to identify the parameters of the consequent. In this paper, we used k-means clustering algorithm to improve the fitness of data description and we used LSM to minimize the inaccuracy when parameters are identified. Detailed explanation will be handled in section 4.

And as we mention in introduction, when predictor uses non-stationary time series to predictor input, prediction performance was not good because of non-linearity of them. Thus, if the mean and dispersion of them are suitably stabilized, it will reveal potential patterns and rules associated with them, appropriately. Thus, to use difference data will be more useful for prediction system. But it is not easy because it should decide appropriate difference intervals which can reveal potential patterns and rules. This problem will be explained in section 3. Following figure (1) is structure of proposed prediction system. It is constructed by multiple models corresponding to difference intervals respectively. In figure (1), detector detects the candidates of difference interval to construct the multiple models. To detect the candidates, we used correlation analysis [9]. Detailed explanation will be handled in section 3.1.

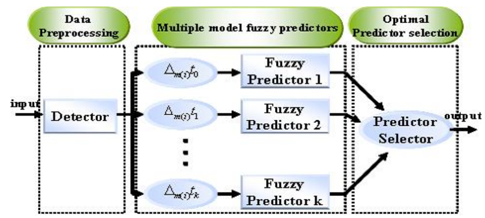


Figure 1) structure of multiple model fuzzy predictors

Also prediction error will be always generated, when each fuzzy predictor performs prediction. If prediction error can be compensated, the accuracy of prediction will be better. Thus, we propose error compensation procedure of using cross-correlation function. This procedure will be explained in section 4.4. In this paper, proposed algorithm is configured by two modes which are training mode and operating mode. Training mode is to construct predictors and operating mode is to perform prediction actually. Procedure of proposed method is as follows and Figure (2) shows flow chart of proposed method.

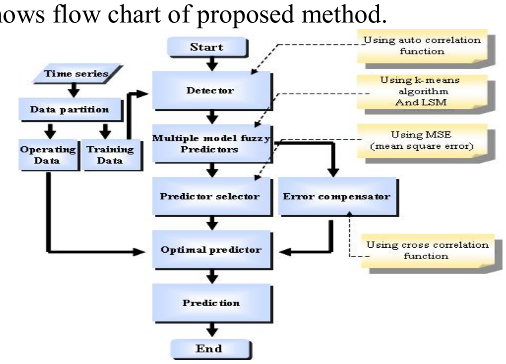


Figure 2) Flow chart of proposed predictors

Training mode

Step 1) Detecting the candidates of difference interval by data preprocessing.

Step 2) Generating the rules of TS fuzzy model corresponding to difference intervals

Step 3) Identifying parameters of TS fuzzy rule using the LSM

Step 4) Performing prediction using training data as Fuzzy model input.

Step 5) Determining error compensation value using cross-correlation coefficient

Step 6) Modifying the consequent of TS fuzzy rule by the error compensation

Step 7) Selecting the best predictor whose MSE is minimum in model bank → the end of design

Operating mode

Step 8) accomplishing prediction using the best predictor

3. Data preprocessing and model selection

Difference data is more useful to predictor that predicts non-stationary time series which include trend such as the periodicity and the increase tendency because it can effectively reveal the potential patterns and rules of them. However, optimum difference intervals will be changed according to the data characteristic. Thus, Data preprocessing is to select optimum differential intervals about this problem. In training mode, detector detects the candidates of difference interval (data preprocessing), and predictors are constructed by using the detected candidates, respectively. Also, in this mode, predictors perform prediction with training data, and then the model selector selects best predictor which minimized the stipulated performance index (model selection). In operating mode, the above selected predictor performs prediction with operating data using optimum difference interval selected from training mode.

3.1 Data preprocessing

Firstly in data preprocessing procedure, training data should be appropriately determined to train predictors, and then auto-correlation coefficients are computed by following formula (1).

$$r_j = \frac{\sum_{i=1}^{N-j}(y(i)-\bar{y})(y(i+j)-\bar{y})}{\sum_{i=1}^{N}(y(i)-\bar{y})^2} \tag{2}$$

Where, N is the length of training data and j is difference interval. Also $y(i)$ is i th training data and $\bar{y}$ is the mean of total training data. The candidates are detected using subtraction between neighbor coefficients after array of computed correlation coefficient in order of the grade. In other words, it means that the intervals corresponding to higher coefficients than coefficient with maximum subtraction become the candidates. And then difference data is generated by the detected candidates $\{m(i)\}$ to use as predictor input, respectively. Difference data at time t can be generated by formula (3). In the formula (3), the number of difference data generated from training data will be changed according to difference intervals $m(i)$.

$$\begin{aligned}
\Delta_{m(i)}t_0 &= y(t) - y(t-m(i)) \\
\Delta_{m(i)}t_1 &= y(t-1) - y(t-1-m(i)) \\
&\vdots \\
\Delta_{m(i)}t_n &= y(t-n) - y(t-n-m(i)) \\
&\vdots \\
\Delta_{m(i)}t_{t-m(i)-1} &= y(m(i)+1) - y(1)
\end{aligned} \tag{3}$$

This is to only generate them using efficient data value and to prevent the bias which is happened by conditions such as y(n)=0, n≤0. So, generated differences are used to their input and parameter identification for each rule base of predictors.

3.2 Model selection

If the number of the candidates is k, the number of predictors will be k too. Thus, k predictors form parallel structure and perform prediction using generated differences respectively. While the above predictors perform prediction with training data, MSE that is defined as follows can be given using predicted result of predictors, respectively.

$$MSE = \frac{1}{N-m(i)} \sum_{n=m(i)+1}^{N} (y(n) - \hat{y}(n))^2 \tag{4}$$

Where, $\hat{y}(n)$ is prediction value of $y(n)$ and N is length of training data. Length of difference data corresponding to difference intervals $m(i)$ is $N-m(i)$. Thus, formula (1) will give unbiased MSE for each predictor. To select best predictor that will perform prediction with operating data, the model selector selects the predictor with minimal MSE because the best predictor will accomplish most good prediction within model bank.

4. Design of TS fuzzy predictor

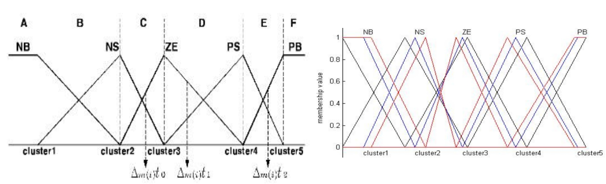
We used TS fuzzy model to perform prediction using differences generated from the above data preprocessing procedure because it can admit advantage of linear regression model and can properly describe linguistic characteristics by small input variables and fuzzy sets. In this paper, we used three difference data such as $\Delta_{m(i)}t_0$, $\Delta_{m(i)}t_1$ and $\Delta_{m(i)}t_2$ as input variables in order to consider effectiveness of predictor construction. Where, $\Delta_{m(i)}t_0$, $\Delta_{m(i)}t_1$ and $\Delta_{m(i)}t_2$ are values generated in order of the most recent from time t, and $m(i)$ is difference interval. Also input space was divided into five fuzzy sets such as NB, NS, ZE, PS and PB according to their, respectively. Thus, linguistic rule of i th TS fuzzy predictor corresponding to difference interval $m(i)$ is formed as follows.

$$\begin{aligned} &R_j : if\ \Delta_{m(i)}t_0\ is\ A_j\ and\ \Delta_{m(i)}t_1\ is\ B_j\ and\ \Delta_{m(i)}t_2\ is\ C_j \\ &\quad Then\ \hat{\nabla}_t^j = a_0^j\Delta_{m(i)}t_0 + a_1^j\Delta_{m(i)}t_1 + a_2^j\Delta_{m(i)}t_2 \end{aligned} \tag{5}$$

Where, output value of the consequent are $\hat{\nabla}_t^j = \hat{y}^j(t+p) - y(t)$. Therefore, output value denotes an increase value from current value $y(t)$ to a p-step ahead prediction value $\hat{y}^j(t+p)$ .In fuzzy rule of formula (5), it is necessary to determine each membership function of input values, and to identify parameters of the consequent. If it is necessary to improve prediction performance of fuzzy predictor whose result was unsatisfactory, it will be effective to increase fuzzy sets and input variables. However, it can bring about the complexity of model structure as well as limitation of selecting the number of input variable and fuzzy sets because this will increase the number of parameter which will be identified. Thus, we selected the number of input variables and fuzzy sets by using empirical method based on analysis results which were simulated according to several time series that is useful to time series prediction.

4.1 fuzzy partition of input space

Fuzzy partition corresponding to input values (such as $\Delta_{m(i)}t_0$, $\Delta_{m(i)}t_1$ and $\Delta_{m(i)}t_2$) is divided as follows. Universe of discourse of fuzzy partition is defined between a minimum value and maximum value among $\Delta_{m(i)}N_0,\cdots,\Delta_{m(i)}N_{N-m(i)-1}$ generated from N training data, and fuzzy partition are divided into five fuzzy sets(such as NB, NS, ZE, PS, PB) using triangle membership function with center values which were computed by k-means clustering algorithm. Its form is as follows.



a) Membership function b) K-means clustering algorithm

Figure 3) Fuzzy sets and K-means clustering algorithm

Figure (3) shows the continual change of membership function by k-means clustering. So, fuzzy partition will be independent for k fuzzy predictors that construct multiple model fuzzy predictors respectively. Therefore, the membership function will be differed according to difference data, respectively. In case of figure (3), the membership function corresponding to input x can get as follows.

In parts of A and F: $\mu_{NB}(x)$ and $\mu_{PB}(x) = 1$

In parts of B, C, D and E:

$$\begin{aligned} \mu_L(x) &= \frac{C_R - x}{C_R - C_L} \\ \mu_R(x) &= \frac{x - C_L}{C_R - C_L} \end{aligned} \tag{6}$$

Where, $\mu_L(x)$ denotes the membership function computed by the left center C_L of the cluster that satisfies input x . Also, $\mu_R(x)$ denotes the membership function computed by the right center C_R of the cluster that satisfies input x . For example, the membership function of figure (3) is as follows.

$$\begin{aligned} &if\ x = \Delta_{m(i)}t_1\ then\ C_L = cluster3, \\ &\quad \mu_L(x) = \mu_{ZE}(x)\ and\ \mu_R(x) = \mu_{PS}(x) \end{aligned} \tag{7}$$

4.2 Fuzzy rule generation

In theory, 125 fuzzy rules can be generated because we used 3 input variables and 5 fuzzy sets. But, if input variables do not satisfy the antecedent of TS fuzzy rule, rule generation is impossible because data that can identify parameters of the consequent of TS fuzzy model do not exist. Thus, this paper generates fuzzy rules when difference data $\Delta_{m(i)}N_0,\cdots,\Delta_{m(i)}N_{N-m(i)-1}$ generated from the training data only satisfied fuzzy sets. It is sufficiently valid to use these rules because we premised that training data are similar to future data, also we used training data to fuzzy partition. Three difference values of neighbor values make one set for the fuzzy sets and rule generation. This set is generated continually when the difference values are slid to one-step with time. Therefore, whole input sets will be given as follows.

$$\begin{aligned} &\{\Delta_{m(i)}N_{N-m(i)+1}, \Delta_{m(i)}N_{N-m(i)}, \Delta_{m(i)}N_{N-m(i)-1}\}, \\ &\{\Delta_{m(i)}N_{N-m(i)+2}, \Delta_{m(i)}N_{N-m(i)+1}, \Delta_{m(i)}N_{N-m(i)}\}, \\ &\qquad \vdots \\ &\{\Delta_{m(i)}N_1, \Delta_{m(i)}N_2, \Delta_{m(i)}N_3\} \end{aligned} \tag{8}$$

Thus, each rule base of predictors is constructed by avoiding same rules that were generated from input data sets respectively. In case of the input set { $\Delta_{m(i)}t_0$,

$\Delta_{m(i)}t_1$, $\Delta_{m(i)}t_2$ } that is showed in figure (3), $\Delta_{m(i)}t_0$ is included in NS and ZE, $\Delta_{m(i)}t_1$ is included in NE and PS, and $\Delta_{m(i)}t_2$ is included in PS and PB of fuzzy sets. Thus, this input set generates eight fuzzy rules as follows.

$$\begin{aligned}
&R_1: if\ \Delta_{m(i)}t_0\ is\ NS\ and\ \Delta_{m(i)}t_1\ is\ ZE\ and\ \Delta_{m(i)}t_3\ is\ PS\ \ then \sim\\
&R_2: if\ \Delta_{m(i)}t_0\ is\ NS\ and\ \Delta_{m(i)}t_1\ is\ ZE\ and\ \Delta_{m(i)}t_3\ is\ PB\ \ then \sim\\
&R_3: if\ \Delta_{m(i)}t_0\ is\ NS\ and\ \Delta_{m(i)}t_1\ is\ PS\ and\ \Delta_{m(i)}t_3\ is\ PS\ \ then \sim\\
&R_4: if\ \Delta_{m(i)}t_0\ is\ NS\ and\ \Delta_{m(i)}t_1\ is\ PS\ and\ \Delta_{m(i)}t_3\ is\ PB\ \ then \sim\\
&\qquad\vdots \qquad\qquad \vdots \qquad\qquad \vdots
\end{aligned} \tag{9}$$

Eliminating fuzzy rules which bring about overlap, all fuzzy rules are generated using all input sets which are generated from the training data, and generated fuzzy rules make the rule base. This rule generation is independently performed to k fuzzy predictors respectively which construct the model bank.

4.3 Parameter identification of fuzzy rule

Parameters of the consequent of TS fuzzy rule can be obtained by the least square method which provides optimal solution using all input sets which satisfy the antecedent of the rules. If input sets which contributed to generate the j th fuzzy rule R_i of the i th TS fuzzy predictor corresponding to differential interval $m(i)$ are n , the linear formula of the consequent become n simultaneous equations as follows.

$$\begin{aligned}
\hat{\nabla}_{i1}^j &= a_0^j\Delta_{m(i)}t1_0^j + a_1^j\Delta_{m(i)}t1_1^j + a_2^j\Delta_{m(i)}t1_2^j\\
\hat{\nabla}_{i2}^j &= a_0^j\Delta_{m(i)}t2_0^j + a_1^j\Delta_{m(i)}t2_1^j + a_2^j\Delta_{m(i)}t2_2^j\\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots\\
\hat{\nabla}_{in}^j &= a_0^j\Delta_{m(i)}tn_0^j + a_1^j\Delta_{m(i)}tn_1^j + a_2^j\Delta_{m(i)}tn_2^j
\end{aligned} \tag{10}$$

These simultaneous equations can be represented to the matrix form as follows.

$$\begin{bmatrix} \hat{\nabla}_p^j(1) \\ \hat{\nabla}_p^j(2) \\ \vdots \\ \hat{\nabla}_p^j(n) \end{bmatrix} = \begin{bmatrix} \Delta_{m(i)}t_0^j(1) & \Delta_{m(i)}t_1^j(1) & \Delta_{m(i)}t_2^j(1) \\ \Delta_{m(i)}t_0^j(2) & \Delta_{m(i)}t_1^j(2) & \Delta_{m(i)}t_2^j(2) \\ \vdots & \vdots & \vdots \\ \Delta_{m(i)}t_0^j(n) & \Delta_{m(i)}t_1^j(n) & \Delta_{m(i)}t_2^j(n) \end{bmatrix} \begin{bmatrix} a_0^j \\ a_1^j \\ a_2^j \end{bmatrix} \tag{11.a}$$

$$Y_j = X_j\Theta_j \tag{11.b}$$

Where, Y_j is output vector, X_j is input data matrix and Θ_j is coefficient vector. Also, the coefficient vector Θ_j of the formula (11.b) can obtain using the least square method as follows.

$$\hat{\Theta}_j = (X_j{}^T X_j)^{-1} X_j{}^T Y_j \tag{12}$$

Presumption coefficients which are obtained by formula will be optimal solution that minimizes the sum of error power such as formula (13).

$$E_j = (Y_j - X\hat{\Theta}_j)^T (Y_j - X\hat{\Theta}_j) \tag{13}$$

4.4 Error compensation

If error which is generated while predictor is performing prediction using the coefficient $\hat{\Theta}^j = [\hat{a}_0^j, \hat{a}_1^j, \hat{a}_2^j]$ that was presumed in parameter identification procedure can be properly compensated, prediction accuracy will be improved. But it is difficult to compensate prediction error because prediction error becomes probability process with characteristic of normal distribution white noise. In this paper, we classified similar pattern data into the homogeneous class (in order words, classifying data sets which satisfy same rule R_j). Thus, we were able to compare individual prediction accuracy of each data that has satisfied same fuzzy rule (R_j). Error compensation for input data is performed appropriately using this procedure and its principle is as follows. Prediction error which was generated from training data with most similar characteristic between an input data set and training data sets in same fuzzy rule R_j is reapplied to error of the input data set for prediction. Although this method is not perfect, it can sufficiently decrease prediction error.

Cross-correlation function is appropriate to compute the similarity between an input set and training data sets in same fuzzy rule R_j . If an input set X_t for prediction is $[\Delta_{m(i)}t_0, \Delta_{m(i)}t_1, \Delta_{m(i)}t_2]$ and a training data set T_n is $[\Delta_{m(i)}n_0, \Delta_{m(i)}n_1, \Delta_{m(i)}n_2]$, cross-correlation coefficient ρ_{XT} is defined as follows.

$$\rho_{XT} = \frac{C_{XT}}{\sqrt{C_{XX}}\sqrt{C_{TT}}} \tag{14}$$

Where, C_{XX} is covariance of X_t , C_{TT} is covariance of T_n and C_{XT} is cross-covariance between X_t and T_n. Also, each covariance is defined as follows.

$$C_{XX} = \sum_{l=0}^{2} (\Delta_{m(i)}t_l - \overline{X}_t)^2 \tag{15}$$

$$C_{TT} = \sum_{l=0}^{2} (\Delta_{m(i)} n_l - \overline{T}_n)^2 \qquad (16)$$

$$C_{XT} = \sum_{l=0}^{2} (\Delta_{m(i)} t_l - \overline{X}_t)(\Delta_{m(i)} n_l - \overline{T}_n) \qquad (17)$$

Where, $\overline{X}_t$ is the mean of X_t and $\overline{T}_n$ is the mean of T_n. Cross-correlation coefficient of the formula (14) has $-1 \le \rho_{XT} \le 1$. If $|\rho_{XT}|$ is 0, it is no relationship, and if $|\rho_{XT}|$ is 1, it denotes perfect linear relationship. Thus, the similarity between two data sets will be higher when cross-correlation coefficient is larger. Therefore, error corresponding to maximum cross-correlation coefficient of errors generated from training sets is applied as compensation error $\hat{e}_t^j$. Finally, the consequent of fuzzy rule R_j for an input set X_t is modified as follows using selected error.

$$\hat{\nabla}_t^j = \hat{a}_0^j \Delta_{m(i)} t_0 + \hat{a}_1^j \Delta_{m(i)} t_1 + \hat{a}_2^j \Delta_{m(i)} t_2 + \hat{e}_t^j \qquad (18)$$

In this procedure, if cross-correlation coefficient is small than given criterion level (we used 0.5), similarity is regarded to zero. Thus, in this case the consequent of fuzzy rule R_j is not modified. Using the fitness μ_j of the antecedent and output $\hat{\nabla}_t^j$ of the consequent in same rule R_j, prediction value $\hat{\nabla}_t$ for an input set X_t of fuzzy predictor with q fuzzy rules can be obtained by the sum of weights as follows.

$$\hat{\nabla}(t) = \frac{\sum_{i=1}^{q} \mu_i \hat{\nabla}_t^i}{\sum_{i=1}^{q} \mu_i}, \quad \hat{y}(t+p) = y(t) + \hat{\nabla}(t) \qquad (19)$$

Prediction value $\hat{\nabla}_t$ of fuzzy predictor is increase value between current value and predicted future value at time t. Therefore, future value $\hat{y}(t+p)$ of original time series is obtained as the formula (19).

5. Simulations and investigation

To verify the effectiveness of proposed techniques, we performed simulations. We used some typical time series with non-stationary and non-linear characteristics to examples of simulations, and then analyzed prediction results. Finally, we compared performance of our method with performance of other research.

In first example, we simulated Australian quarterly electricity production data. [10] This data has 155 data, and we used 70 data to training data and 85 data to operating data. To compare performance, we used MRE (Mean Relative Error) which is defined as follows.

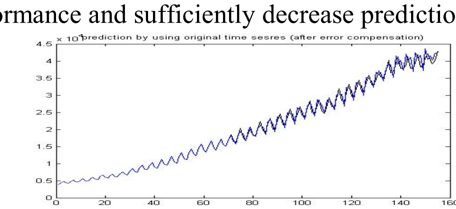
$$MRE = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{|y(i) - \hat{y}(i)|}{|y(i)|} \times 100 \right)\% \qquad (20)$$

Simulations are performed by two cases as follows.

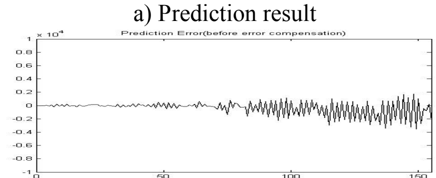
Case 1, predicting by proposed predictors whose structure does not include error compensation procedure by using original data.

Case 2, predicting by proposed predictors whose structure includes error compensation procedure by using difference data.

Figure (4) is prediction performance of case1. In figure (a), the red line is original value and blue line is prediction value. This result shows that prediction is not good. Also, figure (b) shows that case 1 gives considerably large error. Figure (5) shows prediction performance of case 2. These results show that proposed method can give good prediction performance and sufficiently decrease prediction error.

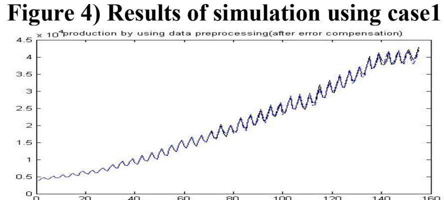


a) Prediction result

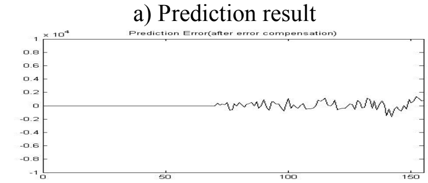


b) Prediction error

Figure 4) Results of simulation using case1



a) Prediction result



b) Prediction error

Figure 5) Results of simulation using case2

Following table (1) shows results that compared each prediction performance of simulations using MRE. In table (1), proposed method whose MRE is 1.6992 provides most good performance. Therefore, these results show that the prediction performance of using difference data of optimal interval generated from data preprocessing is better than that of using original data. Also these results show that the error compensation procedure can compensate the error appropriately when predictors perform prediction.

Table 1) Performance comparison of each case

	case 1		case 2	
	before error compensation	after error compensation	before error compensation	after error compensation
MRE	4.2899	4.1324	2.1720	1.6992
optimal value of differential interval	-		8	8

Following table (2) shows results that compared with the performances of other method [4-6] [11] of using soft computing methods and the performance of our method using same time series, and darkly expressed part is the performance of proposed method. The results in table (2) show that performance of proposed method is more excellent than GA-RS method that was most recently proposed.

Table 2) Performance comparison with traditional method

	Mamdani fuzzy model	multi-fuzzy model	Fuzzy AR	GA-RS method	proposed method
MRE	7.8123	2.7125	3.1254	1.8100	1.6992

In second example, we used Mackey-glass time series that has been frequently using to non-linear time series prediction. Mackey-glass time series is generated by formula defined as follows.

$$\frac{dx(t)}{dt} = \frac{ax(t-\tau)}{1+x^{10}(t-\tau)} - bx(t) \tag{21}$$

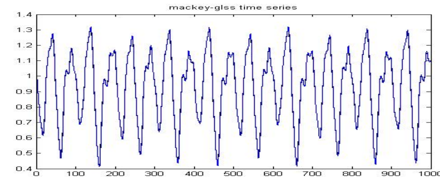
Parameters of this formula were predefined by a =0.2, b =0.1 and τ =17. Whole simulation data are selected 1000 data from $x(124)$ to $x(1123)$. We used 500 data of total data to training data, and remainder data to operating data.

Simulations were performed on two cases.

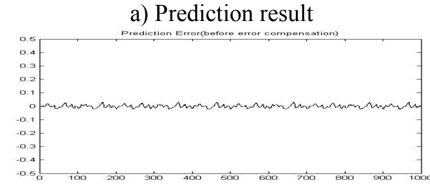
Case 1, predicting by proposed predictors whose structure does not include the error compensation mechanism by using difference data.

Case 2, predicting by proposed predictors whose structure includes the error compensation procedure by using difference data.

Figure (6) shows result of case 1. In figure (a), black line is original value and blue line is prediction value. Figure (b) shows prediction error. These results show that difference data is more effective to predict non-linear time series.



a) Prediction result

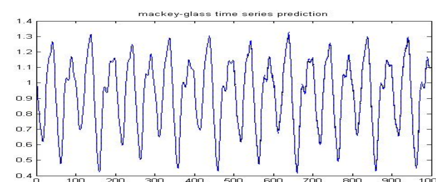


a) Prediction error

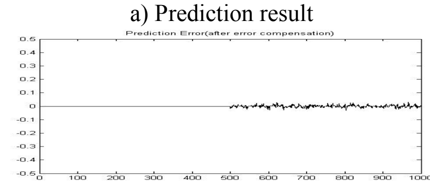
Figure 6) Results of simulation using case1

Figure (7) shows result of case 2 that used error compensation procedure. We used RMSE (Root Mean Square Error) defined as follows to performance index for this time series.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y(i)-\hat{y}(i))^2} \tag{22}$$



a) Prediction result



a) Prediction error

Figure 7) Results of simulation using case2

RMSE is 0.0122 in case1, and 0.0085 in case2. This result shows that error compensation procedure can decrease prediction error more than case 1. Following table 3 are results that compared performance of our method with Wang' method [12], MCM-1 and MCM-2[13]. Table (3) shows that prediction performance of proposed method is better than others. Also table (3) shows that proposed method performed excellent prediction using smaller fuzzy partition.

Table 3) Performance comparison with traditional method

the number of fuzzy set	RMSE			
	WANG	MCM-1	MCM-2	our method
5	-	-	-	0.0085
7	0.0372	0.0374	0.0340	-
11	0.0253	0.0265	0.0235	-
15	0.0191	0.0197	0.0187	-
19	0.0161	0.0162	0.0159	-
23	0.0135	0.0142	0.0131	-
27	0.0115	0.0115	0.0113	-
31	0.0108	0.0108	0.0105	-

Therefore, these results of simulations show that prediction performance can be improved by proposed method that used difference data that can sufficiently reveal non-stationary time series as input of predictor. Also results of these simulations show that prediction performance can be more improved using error compensation procedure.

6. Conclusion

In time series prediction fields, predicting non-stationary and/or chaotic time series whose long-history has the drift and non-linearity is not easy. Using simulations on some typical examples, in this paper, we showed that using difference data as predictor input is better than original data on non-stationary time series prediction. These results can mean that our method that used difference data reveals patterns and rules in non-stationary time series better. Thus, to generate difference data, we adopted data preprocessing based on correlation analysis. Through this procedure, the candidates of difference interval for generating difference data are determined, and then difference data are generated by using the candidates. Secondly, we applied k-means clustering algorithm and LSM to minimize the inaccuracy that can be produced during predictor modeling. These methods were able to perform more effectively fuzzy partition and parameter identification for modeled predictors. Also among the predictors that are modeled by using the candidates, predictor with minimum performance index during training mode was selected in order to predict future behavior. These procedures gave the good result since selected predictor can predict better than others. Finally, to improve the prediction accuracy more, we added error compensation procedure based on correlation analysis and we were able to get higher prediction accuracy. Thus, we think that proposed method will be applied to field of non-stationary time series prediction usefully.

In future, adaptation techniques that can select predictor corresponding to considerable change of input data with time will need to research. Also the error compensation method that can decrease the complexity of proposed method and effectively compensate error will need to research.

7. References

[1] George E. P. Box and Gwilym M. Jenkins, *Time series analysis: Forecasting and Control*, Holden-Day, 1970.

[2] George J. Klir and Bo Yuan, *Fuzzy Sets and Fuzzy Logic Theory and Applications*, Prentice-Hall, 1995

[3] Juhong Nie, "Nonlinear Time-Series Forecasting: A Fuzzy Neural Approach", *Neuro computing*, MacMaster University, vol.16, 1999, pp.66-76,

[4] K.Ozawa, T.Niimura, "Fuzzy Time-Series Model of Electric Power Consumption", *IEEE Canadian Conference on Electrical and Computer Engineering*, 1999, pp.1195-1198

[5] Daijin Kim, Chulhyun Kim, "Forecasting Time Series with Genetic Fuzzy Predictor Ensemble", *IEEE Trans. on Fuzzy Systems*, 1997, vol. 5, pp.523-535

[6] Inteak Kim, Song-Rock Lee, "A Fuzzy Time Series Prediction Method based on Consecutive Values", *IEEE International Fuzzy Systems*, 1999, vol.2, pp.703-707

[7] Chul-Heui Lee, Sang-Hun Yoon, "Fuzzy Nonlinear Time Series Forecasting with Data Preprocessing and Model Selection", *Journal of Telecommunications and Information*, 2001, vol.5, pp.232-238

[8] Carl G. Looney, "Pattern Recognition using Neural Networks", *Oxford University Press*, 1997

[9] Haekyoung Kim, Taesu Kim, *time series analysis and prediction theory*, munundang in Korea, 2003

[10]http://www.personal.buseco.monash.edu.au/hyndman/TSDL/

[11] Youngsuk Joo, *Fuzzy System Modeling using Genetic Algorithm and Rough Set Theory*, master' degree thesis, kangwon national university in Korea, 2003

[12] L. X. Wang, J. M. Mendel, "Generating fuzzy rules from numerical data, with applications", *IEEE Trans. on Systems*, Man, and Cybern, 1992, No.6, pp1414-1427

[13] Intaek Kim, Changwook Kong "Fuzzy Learning Algorithms for Time Series Prediction", *Journal of intelligence and Information systems,* 1997, vol.7, No.3, pp34-42