2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

Fuzzy Model Identification for Time Series System Using Wavelet Transform and Genetic DNA Code

Yeun-Woo Lee*, Jung-Chan Kim* and Young-Hoon Joo*

* School of Electronic & Information Engineering, Kunsan National University, Kunsan, Jeonbuk 573-701, Korea

E-mail: yhjoo@kunsan.ac.kr

***Abstract–* In this paper, we propose a new fuzzy model identification of time series system using wavelet transform and genetic DNA code. Generally, it is well known that the DNA coding method is more diverse in the knowledge expression and better in the optimization performance than the genetic algorithm (GA) because it can encode more plentiful genetic information based on the biological DNA. The proposed method can construct a fuzzy model using the wavelet transform, in which the coefficients are identified by the DNA coding method. Thus, we can effectively get the fuzzy model of the nonlinear system by using the advantages of both wavelet transform and DNA coding method. In order to demonstrate the superiority of the proposed method, it is compared with modeling method using the conventional GA.**

I. INTRODUCTION

When researcher wants to find the model of a system mathematically, the differential equation has been widely used. However, there are so much nonlinearity and a number of time constants in realistic system that the accurate differential equation can hardly be obtained. Though comparatively precise model is acquired, the efficiency decreases by model approximation. In order to solve this problem, the fuzzy inference system introduced by Zadeh.[1].

The conventional fuzzy model is based on the knowledge of an expert, and the parameters and the structure of a fuzzy model are tuned through trial and error. But it is time-consuming. Recently, many self-tuning methods such as the genetic algorithm(GA) by which the parameters and the structure of a fuzzy model are tuned have been studied.[2]. But conventional GAs encode solution space to the fixed position, fixed length strings. Also, it is difficult to acquire properly linkages associated with a given problem because of encoded solution structures are not known. So, this weak linkage mean that building-block is likely to breakdown. To prevent this problem, Joo proposed a hybrid algorithm using both the GA and the fuzzy c-means clustering, and a fuzzy modeling using the messy genetic algorithm (mGA) [3]. And recently, the DNA coding method was studied to solve the problem of GA [4].

The DNA coding method is more diverse in the knowledge expression and better in the optimization performance than the GA because it can encodes more plentiful genetic information based on the biological DNA [4]. A living body starts its life from the DNA, the primary information career in genetics. Almost every critical activity of the organism is accomplished by proteins constructed from the DNA. The efficacy of the organism, i.e, the genetic fitness, depends on the proteins. For some reason our body chooses different representations of the information stored in the proteins. It uses the mRNA and the DNA sequences to represent the proteins.

On the other hand, Wang and Zeng [5] represented the fuzzy system with the linear combination of the fuzzy basis function (FBF), and Lin [6] obtained the equivalent model to the discrete wavelet transform through the modification of the fuzzy model. Hereby, the problem that the conventional fuzzy model can hardly deal with the abrupt change of signal was resolved. The wavelet transform was very effective in analyzing the physical status of a certain signal with singularity and with an element of high frequency as compared with the Fourier transform. The wavelet theory has been developed by Daubechies [8], and made rapid progress by proof of Donoho [7] that wavelet can be a basis function for any signal using unconditional basis. The fuzzy system modeling based on Donoho's unconditional basis has the advantage of wavelet transform by constituting FBF and consequent part to equalize the linear combination of FBF with the linear combination of wavelet function and modifying fuzzy system model to be equivalent to wavelet transform. In this paper, we propose a new fuzzy model identification of time series system using wavelet transform and genetic DNA code. The proposed method is applied to fuzzy modeling of Box-Jenkins time series system.

II. WAVELET TRANSFORM AND FRAME

Any function or signal in $L^2(R)$ can be represented as the linear combination of basis function, in the following form[6];

$$f(x)=\sum_k C_{j_0}(k)\varphi_{j_0,k}(x)+\sum_k\sum_{j=j_0}^{\infty} d_j(k)\psi_{j,k}(x) \qquad (1)$$

where, $\varphi_{j_0k}(x)$ is scale function and $\psi_{j,k}(x)$ is the wavelet function. In Eq. (1), the first sum is approximation of low resolution and the second sum is approximation of high resolution. The wavelet transform is the mapping of $L^2(R)\rightarrow L^2(R)^2$ and is separated into a low band pass filter, scale function and a high band pass filter, wavelet function. If $\varphi_{j_0k}(x)$ and $\psi_{j,k}(x)$ is orthogonal each other for any j and k the coefficients of Eq.(1) $c_j(k)$ and $d_j(k)$ can be obtained by inner product. Finding the coefficients is called wavelet transform.

The Eq.(1) is elucidated by multi-resolution analysis. The various subspaces can be seen from the following expression. As illustrated in Fig.1, at a given $j=-\infty$, the following equation can be considered. The multi-resolution equation can represent the signal with time-scale by dividing into precise elements.

$$L^2 = \cdots \oplus W_{-2} \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus W_2 \cdots \quad (2)$$

therefore, $\psi_{j,k}(x)$ of Eq.(1) can represented by,

$$\psi_{j,k}(x) = a^{-j/2}\psi(a^{-j}x - bk) \quad (3)$$

A multidimensional wavelet function is represented with tensor product of single dimensional wavelet function as follows;

$$\psi(x) = \psi_1(x_1)\cdots\psi_n(x_n) \quad (4)$$

Assuming that single dimensional wavelet transform is separated into n orthogonal direction elements, Fourier transform of each term in Eq.(4) is substituted for itself.

$$\hat{\psi}(x) = \hat{\psi}_1(x_1)\cdots\hat{\psi}_n(x_n) \quad (5)$$

where, $\hat{\psi}(w)$ is Fourier transform of $\psi(w)$.

$$\int \frac{|\hat{\psi}_i(\omega_i)|^2}{|\omega_i|} d\omega_i \langle \infty \quad (6)$$

$$A\|f\|^2 \leq \sum_{j,k} |\langle f, \psi_{j,k} \rangle|^2 \leq B\|f\|^2 \quad (7)$$

where $A \rangle 0, B \langle \infty$.

Therefore, $\psi_i(x_i)$ which satisfies conditions (6) and (7) should be set as wavelet frame. In this thesis, 'Mexican Hat' is employed as a mother wavelet function which satisfies both of the conditions as follows:

$$\psi_i(x_i) = \alpha_i(1 - \alpha_i x_i^2)e^{-\frac{\alpha_i x_i^2}{2}} \quad (8)$$

'Mexican Hat' in Eq. (8) is derived from the function which is in proportion to the second-ordered differential form of the Gaussian probability density function.

Substituting Eqn. (8) for (3) and (4), the following is computed

$$\psi_{j,k} = a^{-\frac{j_1}{2}}\alpha_1[1 - \alpha_1(a^{-j_1}x_1 - b_1k_1)^2]e^{-\frac{\alpha_1(a^{-j_1}x_1 - b_1k_1)^2}{2}} \cdots$$

$$a^{\frac{j_n}{2}}\alpha_n[1 - \alpha_n(a^{-j_n}x_n - b_nk_n)^2]e^{-\frac{\alpha_n(a^{-j_n}x_n - b_nk_n)^2}{2}}. \quad (9)$$

III. FUZZY MODEL AND FBFs

The fuzzy model is modified so that it can be equivalent to discrete wavelet transform. The fuzzy rule is to represent the characteristic of a system linguistically. The fuzzy model which is used in this thesis is as follows;

$$\textit{Rule } i: \textit{If } x_1 \textit{ is } A_{i1}, \cdots, x_n \textit{ is } A_{in},$$
$$\textit{Then } y_i \textit{ is } d_i\alpha_1(1-\alpha_1 x_1^2) \cdots \alpha_n(1-\alpha_n x_n^2) \quad (10)$$

where Rule i is the ***i-th*** rule, x_j is the ***j-th*** input variable, y_i is ***I-th*** output variable, and A_{ij} is a membership function for the ***i-th*** rule of the ***j-th*** input defined as a Gaussian function. The conclusion part is constituted by the product of the remainders except Gaussian function of 'Mexican Hat' wavelet function in Eq. (9).

That is the model which equivalent to wavelet transform can be obtained by modifying the consequent part of the general fuzzy model. The output for the arbitrary fuzzy rule basis in the fuzzy inference system which has several fuzzy rule bases can be represented with the linear combination of FBFs in the following;

$$y_i = \sum_{i=1}^{c} B_i(x)y_i \quad (11)$$

Substituting Eq. (10) for Eq. (11) yields the followings:

$$y_i = \sum_{i=1}^{c} B_i d_i \; a^{-\frac{j_1}{2}}\alpha_1(1 - \alpha_1(a^{-j_1}x_1 - b_1k_1)^2)\cdots$$
$$a^{-\frac{j_1}{2}}\alpha_1(1 - \alpha_1(a^{-j_1}x_1 - b_1k_1)^2), \quad (12)$$

where $B_i = \frac{\prod_{k=1}^{n} A_{ik}(x_k)}{\sum_{i=1}^{c}\prod_{k=1}^{n} A_{ik}(x_k)}, d_i \in R.$

Therefore, the final output y can be obtained as follow:

$$y = \sum_{j=1}^{N} y_i = \sum_{j=1}^{N}\sum_{k=1}^{c} d_{j,k}\psi_{j,k} \quad (13)$$

Equation (13) means that the fuzzy model is equivalent to discrete wavelet transform through modifying the conclusion part of a fuzzy model which has several fuzzy rule bases. Using 'Mexican hat' wavelet function, the Gaussian function part in the wavelet function becomes membership function of the fuzzy model, $A_{ik}(x_k)$ and the product of the parts except the Gaussian function becomes the conclusion part of the modified fuzzy model.

IV. DNA CODING METHOD

The DNA coding method is optimization technique based on biological DNA structure. It is more diverse in the knowledge expression and better in the optimization performance than the GAs because it can encodes more plentiful genetic information[4].

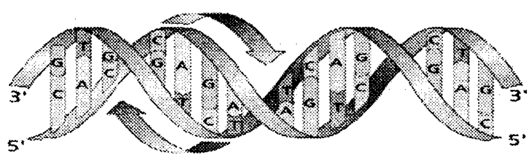


Fig.1 DNA Double Helices

The DNA is the primary carrier of the genetic information that is transmitted from one generation to another, and DNA molecules consist of two long complementary chains held together by base pairs. DNA consists of four kinds of bases joined to a sugar-phosphate backbone. The four bases in DNA are Adenine *(A)*, Guanine *(G)*, Thymine *(T)* and Cytosine *(C)*. Chromosomes are made of DNA double helices. Where, bases in DNA helices obey the complementary base pairing rule. T and G pair with A and C respectively.

The information coded in the DNA is extracted during the process of gene expression and expression of genetic information coded in DNA requires construction of the mRNA sequence, followed by that of proteins. Also, DNA is decaled by mRNA and is translated by protein in ribosome. Fig. 2 shows the above process.

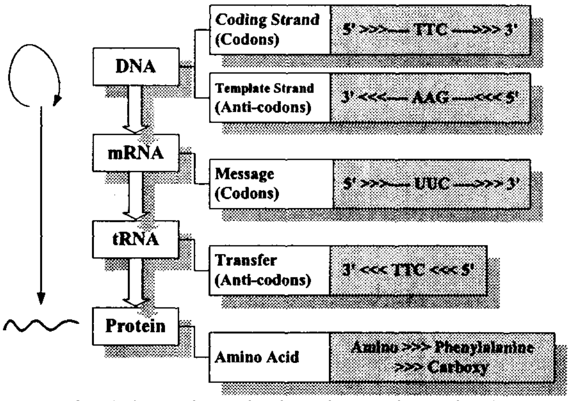


Fig. 2 Strands and Directions of Synthesis

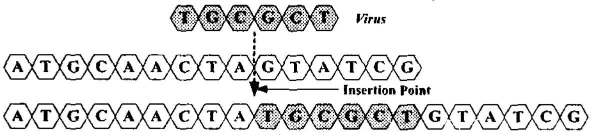
Basic urea of biologic DNA is the four nucleotide. Nucleotide is consisted of A, G, C, and T. It is comprised of three adjacent nucleotides in DNA and forms one codon. Such 64 codons are translated by amino acid which have genetic information.

Note. Number of The Amino Acid

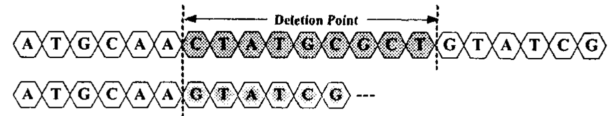
a) 1 codon: ***Met, Trp*** (Start codon)
b) 2 codon: Asn, Asp, Cys, Gln, Glu, His, Lys, Phe, Tyr
c) 3 codons: Ile, ***Stop***(*"nonsense"*)
d) 4 codons: Ala, Gly, Pro, Thr, val
e) 5 codons: none
f) 6 codons: Arg, Leu, Ser

Start codon : ATG, TGG / ***Stop codon*** : TAA, TAG, TGA

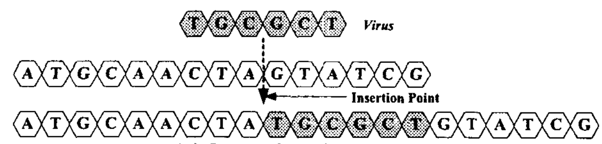
Therefore, codon is translated into amino acid. According to genetic information of amino acid, fuzzy rule is created. Also, each individual evolve using the genetic operators-crossover, mutation, deletion, insertion, and inversion. Individual with high fitness is reproduced in next generation via these process. Fig. 3 shows arithmetic principle of each genetic operator, and Fig. 4 shows genetic operators proposed in this paper,



(a) Deletion Operator



(b) Insertion Operator



(c) Inversion Operator

Fig. 3 Genetic Operators

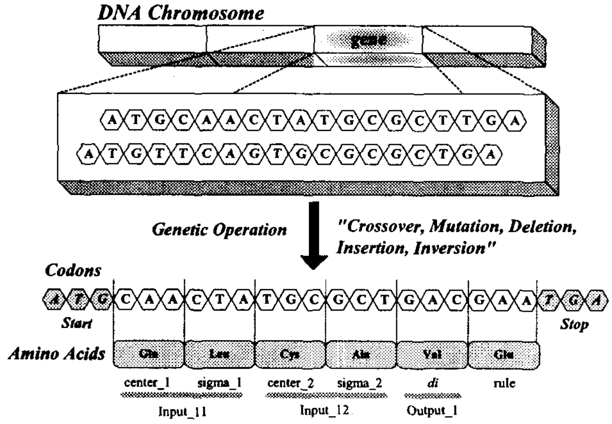


Fig. 4 The Structure of a String

V. EMPIRICAL RESULTS

We consider the fuzzy modeling of a dynamic process using a well-known ezample proposed by Box and Jenkins[10]. The process is a gas furnace with input $u(t)$ and output $y(t)$, which are gas flow rate and CO_2 concentration, respectively.296 I/O data pairs are used in this example. Since the process is dynamic, we take $u(t-4)$ and $y(t-1)$ as input variables to model the fuzzy inference system. All training data pairs are normalized to between 0 and 1. The initial parameters used in this thesis is show in Table 1.

Table 1. Initial Parameter set

Parameters	Value
Generation number	300
Population number	400
Maximum rule number	10
Crossover rate	0.9
Mutation rate	0.01
Deletion rate	0.01
Insertion rate	0.01
Inversion rate	0.01

Also, Fig. 5 shows that the MSE is effectively decreased in early stage. Fig. 6 shows changes of the number of rules. Fig. 7 compares the actual output of the gas furnace and output from the identified fuzzy model. As shown in this figure, the proposed method modeled the actual output approximately. In Table 2., we compare the performance of our fuzzy model with other ones.

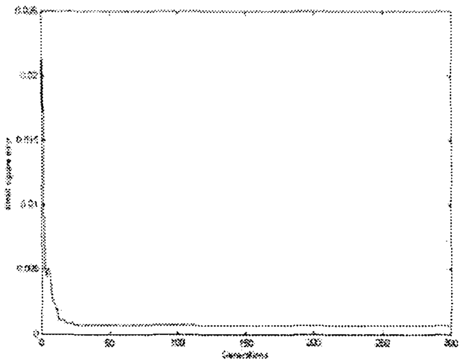


Fig. 5 Change of MSE

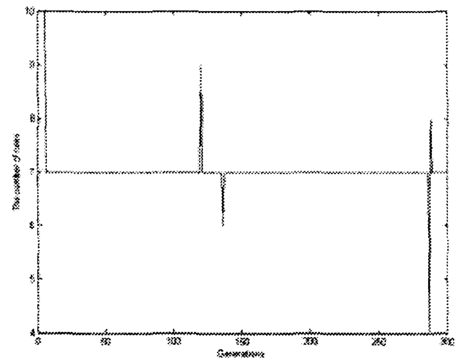


Fig. 6 Change of the number of rules

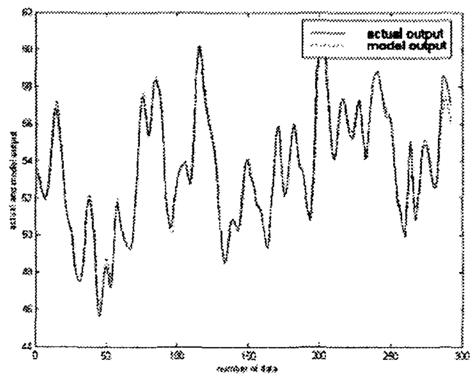


Fig. 7 Identification Result

Table 2. Comparison between our model and other models

Model	Inputs	Rules	MSE
Tong's[10]	y_{k-1}, u_{k-4}	19	0.685
Pedrycz's[11]	y_{k-1}, u_{k-4}	81	0.566
Xu's[12]	y_{k-1}, u_{k-4}	25	0.573
Liska's[13]	y_{k-1}, u_{k-4}	10	0.367
Our's	y_{k-1}, u_{k-4}	7	0.0007

VI. CONCLUSION

In this paper, we proposed a fuzzy model identification of time series system using wavelet transform and genetic DNA code. Modifying conventional fuzzy model to have several fuzzy rule bases and the consequent part to be the part of the 'Mexican Hat' wavelet function, the fuzzy model was equivalent to discrete wavelet function. The wavelet-based fuzzy model, thus, inherited the advantage from discrete wavelet transform. When an arbitrary signal is represented with linear combinations of wavelet functions, the energy can be compacted. It was confirmed that the prominent fuzzy model can be obtained with the small number of rules by this property. In order to, demonstrate the superiority and efficiency of the proposed scheme, we applied this method to the fuzzy modeling of the Box-Jenkins time series system.

This work has been supported by EESRI(R-2003-B-078, which is funded by MOCIE(Ministry of commerce, industry and energy.

REFERENCES

[1] L. A. Zadeh, "Fuzzy Sets", Int, Jour. of Information control, vol. 8, pp. 338-353, 1965.
[2] Y. H. Joo, H. S. Hwang, K. B. Kim, and K. B. Woo, "Linguistic Model Identification for Fuzzy System" Electronics Letters, Vol. 31, No. 4, pp. 330-331, 1995.
[3] Y. H. Joo, Y. W. Lee, J. B. Park, "mGA Based Fuzzy Model Identification", Electronics Letters, vol. 49D, No. 8, pp. 407-414, 2000. 8.
[4] T. Yoshikawa, T. Furuhashi, Y. Uchikawa, "A Combination of DNA Coding Method with Pseudo-Bacterial GA for Acquisition of Fuzzy Control Rules", The 1st Online Workshop on Soft Computing(WCSI), 1996. 8
[5] X. J. Zeng and M. G. Singh, "Approximation Accuracy Analysis of Fuzzy Systems with the Center-Average Defuzzifier", Proc. FUZZ-IEEE/IFES'95, pp. 109-116, 1995.
[6] C. K. Lin and S. D. Wang, "Fuzzy Modeling Using Wavelet Transforms", Electronics Letter, vol. 32, pp. 2255-2256, 1996.
[7] A. Bruce, D. Donoho, H. Y. Gao, "Wavelet Analysis for Signal Processing", IEEE Spectrum, vol, 33, pp. 26-35, 1996
[8] I. Daubechies, "The Wavelet Transform, Time-frequency Localization and Signal analysis," IEEE Trans. on Inform. Theory, Vol. 36, pp. 961-1005, 1990.
[9] G. E. P. Box and G. M. Jenkins, Time Series Analysis, Forecasting and Control, Holden Day, 1970.
[10] R. M. Tong, "The Evaluation of Fuzzy Models Derived from Experimental Data", Fuzzy Sets and Systems, Vol. 4, pp. 1-12, 1980.
[11] W. Pedrycz, "An Identification Algorithm in Fuzzy Relational Systems", Fuzzy Sets and Systems, Vol. 13, pp. 153-167, 1984.
[12] C. W. Xu, "Fuzzy Model Identification and Self-Learning for Dynamic Systems", IEEE Trans. On Syst. Man. Cybern., Vol. 17, pp. 683-689, 1987.
[13] J. Liska and S. S. Melsheimer, "Complete Design of Fuzzy Logic Systems Using Genetic Algorithms", Proc. Of the third IEEE Conf. Fuzzy Systems, pp. 1277-1382, 1994.