대한산업공학회/한국경영과학회 2002 춘계공동학술대회
한국과학기술원(KAIST) 2002 년 5월 3 일-4 일

A Multi-Stage TSK Fuzzy Modeling Method by Genetic Programming

Bo Li [1),2)] and Kyu-Kab Cho [1)]
[1)]Department of Industrial Engineering, Pusan National University, Pusan, Korea
[2)]School of Management, Tianjin University, Tianjin, China

Abstract: This paper deals with a multi-stage TSK fuzzy modeling method by using Genetic Programming (GP). Based on the time sequence of sampling data, the best structural change points of complex systems are determined by using GP, and also the moving window is simultaneously introduced to overcome the excessive amount of calculation during the generating procedure of GP tree. Therefore, a multi-stage TSK fuzzy model that attempts to represent a complex problem by decomposing it into multi-stage sub-problems is addressed and its learning algorithm is proposed based on the Radial Basis Function (RBF) network. This approach allows us to determine the model structure and parameters by stages so that the problems can be simplified.

Key Words: Multi-stage TSK fuzzy modeling, recursive genetic programming, the moving window, curse of dimension

1. Introduction

One of the key problems for dealing with the complex systems modeling via fuzzy logic is "Curse of Dimension". With the increase of data samples, the pattern space required for search is becoming so huge that the explosion of combination can occur and be led to search for a lot of meaningless patterns. At the same time, the relationship among the variables is getting more and more complicated. There have been several attempts to solve this problem. Lin and Lee[1] have proposed a rule deletion mechanism to reduce the number of fuzzy rules, but the completeness of the resulted rule set can not be ensured. In Jang's ANFIS model[2], only the two most important inputs are considered and all the others are ignored. It is only rough approximation to the system. A fuzzy modeling method based on the RBF have proposed by Li and Zhang[3] which may be viewed as an improvement to the Jang's method and a cursive learning algorithm has been developed to overcome the shortcoming of time consuming problem. In addition, the other approaches to solve dimensionality problem have been proposed, such as Multi-stage (hierarchical) fuzzy model by Raju[4,5] for feed-water flow control to a steam generator. Its idea is that the whole system is divided into several stages and each stage corresponds to an individual fuzzy reasoning stage, and it has been proved that the number of rules of the multi-stage fuzzy model is approximately the linear function of the input number of the system. Therefore, multi-stage model can effectively reduce the complexity of the systems.

In the real applications, we often encounter the problems that the time sequence relationship among sampling data is a mainly influential relationship, for example, the supervised data in the industrial control, and the finance data in the economic systems. Concerning the fuzzy modeling based on a great amount of time sequence data, we can find out the relationship among the system variables based on the characterization of the time sequence to get rid of the high-dimensional problem. First, we will find out the structural change points of the systems, that is, the huge change about the dynamic behavior of the systems can be led to the change of the system structure patterns. Based on this idea, we can extract the huge change points representing the system developing behavior from the time sequence sampling data and then divide the data set interval into several sub-intervals to build each sub-models. The final model is the combination of these sub-models that allow the consequence of a sub-model passed to another as a fact through the intermediate variables.

In this paper, a multi-stage TSK fuzzy modeling methods based on the time sequence data is proposed. In section two, both types of single-stage and multi-stage TSK fuzzy models will be reviewed. The determination of the best structural change points with the aid of the combination of GP and the moving window is introduced in section three. Concerning the learning algorithm, the procedure based on RBF is summarized in section four, and simulation results and the conclusion are reported in sections five and six, respectively.

2. Single-stage and Multi-stage TSK Fuzzy Model

2.1 Single-stage TSK Fuzzy Model

The Takagi-Sugeno-Kang's fuzzy models have been proved to be effective in complex system modeling[6]. It needs less rules and each rule's consequence with linear function can describe the high nonlinear input-output mapping relation. Recently, in order to simplify the interdependent procedure for the structure determination and parameter identification, a Modified TSK fuzzy model[7] was presented. It views fuzzy models as a class of local modeling approaches which attempt to solve a complex system by decomposing it into a number of simpler sub-problems. The form of this TSK fuzzy model is as follows:

Rule i (R^i): IF X is A^i , THEN

$$y^i = f^i(X) = b_0^i + b_1^i x_1 + \cdots + b_n^i x_n \quad (1)$$

$$i = 1, \cdots, m$$

where,

① $X = [x_1, x_2, \cdots, x_n]^T$ is the n dimension input variables,

② $A^i = [A_1^i, \cdots, A_n^i]^T$ is the n dimension fuzzy subset with membership degree $\mu_{A^i}(X)$ having the N-dimension tuning parameters called the premise parameters,

③ y^i is the output variable, and

④ $b_j^i, j = 0,1,\cdots,n$ is the consequent parameters.

For any input X, the total output y can be obtained by the following fuzzy reasoning :

$$y = \sum_{i=1}^{m} \mu_{A^i}(X) y^i(X) = \sum_{i=1}^{m} \mu_A(X)(\sum_{j=1}^{n} b_j^i x_j + b_0^i)$$

$$\text{and} \quad \sum_{i=1}^{m} \mu_{A^i}(X) = 1 \quad (2)$$

Thus, each rule of the TSK fuzzy model can be regarded as a local model. All of the rules are used to represent the pattern space of complex system. Because all rules have the same premise structure, we may only deal with the identification of premise parameters (i.e., the centers of the rule premises) and consequence parameters during the fuzzy modeling procedure.

2.2. Multi-stage TSK fuzzy model

Given the time sequence set of sampling data $\{x(t), t = 1, \cdots, N\}$, where N is the number of the sampling data, suppose that we put the N data into k stages, that is, there are M_1 data in the first stage and the input dimension is n_1, M_2 data in the second stage and the input dimension is n_2 and so on, as shown in Fig. 1 and $N = \sum_{i=1}^{k} M_i$. Then the multi-stage TSK fuzzy model can be expressed as follows:

Stage 1 R^l: IF X_1 is A_1^l

$$\text{THEN } y_1^l = h_1^l(X_1) \qquad (3)$$

where,

① $X_1 = [x_1, x_2, \cdots, x_{n_1}]^T$ is the n_1 dimension input variables,

② $A_1^l = [A_{11}^l, \cdots, A_{1n_1}^l]^T$ is the n_1 dimension fuzzy subset with membership function $\mu_{A_1^l}(X_1)$,

③ $h_1^l(X_1)$ is the linear function of the inputs,

④ $l = 1, \ldots, m_1$ is the number of the rules.

Given $X_1 = (x_1, \cdots, x_{n_1})^T \in R^{n_1}$, then the output of the stage 1 fuzzy model is calculated by

$$y_1 = \sum_{l=1}^{m_1} \mu_{A_1^l}(X_1) y_1^l(X_1) \qquad (4)$$

Similarly, The form of the i-th stage fuzzy model can be written as:

Stage i : R^l : IF X_i is A_i^l, y_{i-1} is B_{i-1}

$$\text{THEN } y_i^l = h_i^l(X_i, y_{i-1}) \qquad (5)$$

where,

① $X_i = [x_{N_i+1}, \cdots, x_{N_i+n_i}]^T$,

② $N_i = \sum_{j=1}^{i-1} n_j \le n$, n is dimension of the whole system,

③ $A_i^l = [A_{iN_i+1}^l, \cdots, A_{iN_i+n_i}^l, B_{i-1}]^T$ is the $n_i + 1$ dimension fuzzy subset with the membership function $\mu_{A_i^l}(X_i, y_{i-1})$, and

④ $l = 1, \ldots, m_i$ is the number of the rules.

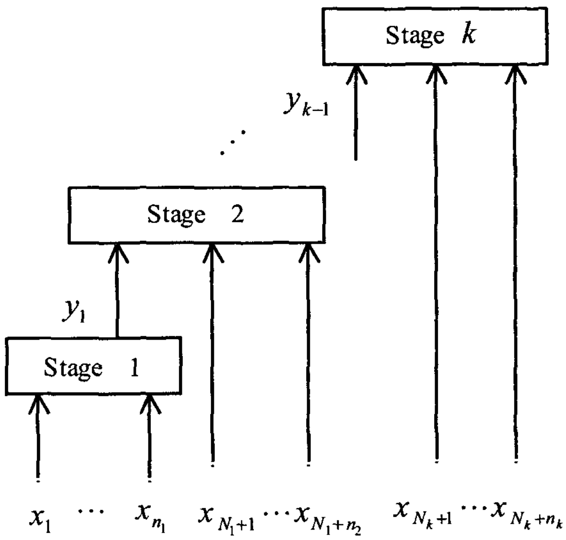

Fig. 1 Multi-stage fuzzy model structure

The output of the i-th stages fuzzy model is as follows:

$$y_i = \sum_{l=1}^{m_i} h_i^l(X_i, y_{i-1}) \mu_{A_i^l}(X_i, y_{i-1}) \qquad (6)$$

Therefore, put M_1 sampling data(n_1

dimension inputs) to the TSK fuzzy model in the first stage and deduce the output y_1, and then another M_2 data(n_2 dimension inputs) and output y_1 in the first stage are sent to the second stages to obtain the output y_2 and so on. The final model output is the output of the k -th stage.

Concerning the above multi-stage TSK fuzzy model, Wang[8] has proved its universal approximation property based on the constructed method.

3. The Multi-stage Structure Determination by using GP

3.1 The Moving Window

Given the data set $\{x(t), t=1,\cdots,N\}$, suppose that N_1 is the width of the moving window and the length of each moving step from left to right is N_2, then the modeling analysis will be held in $s=(N-N_1)/N_2$ moving windows. Because of $N_1 \langle\langle N$, the method based on the moving window decomposes the whole system into several parts according to the time sequence relationship in order to simplify the calculation procedure.

3.2 The Determination of the Structural Change Points by GP

GP proposed by Koza[9] is concerned with the automatic generation of computer programs. Recently, it has grown rapidly and has been applied to a variety of different fields, such as data mining, robot control, patterns recognition, planning and game-playing strategies.

Most GP systems employ a tree structure to represent the program that calculates a function or executes an algorithm. Suppose that the t -th generation of GP is

$$\{f_1^t(X), f_2^t(X), \cdots, f_m^t(X)\}$$

where,

① m is the number of the t -th generation population,

② X is the system we want to study and known the time sequence $X=\{x(t), t=1,\cdots,N\}$, and

③ f is a function which generates from the fundamental operators.

Let us assume the set of functions is:

$$F=\{+,-,\times,\div,\wedge(power), \sin, \cos, \log, \exp, \cdots\} \quad (7)$$

The sampling data in the k -th moving window is called as the terminal set, that is,

$$T_k=\{x(k\times N_2+1),\cdots,x(k\times N_2+N_1\} \quad (8)$$

The trees in GP are composed of the functions and the terminal symbols. The idea used to product every individual in GP is as follows : First, take some function f randomly from the function set as the root node of the GP tree and determine the number of dependent variables according to the function f, for example, "$*$" is an operator which asks for two successor nodes. To every successor node, the offspring node will be chosen from the union set $S=F\bigcup T_k$ of the function set and the terminal set according to some random distribution. The selected function symbols refer to the non-leaf nodes in the GP tree and compute or process an argument passed from its children nodes. However, the terminal symbols correspond to nodes in the leaves of the tree and this branch can't grow up. This procedure is done from up to down and from left to right until the whole tree is finished. An example of the GP tree is shown in Fig. 2.

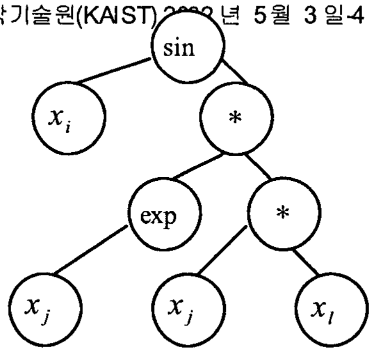

Fig.2 An example of the GP tree

Suppose that the functional form of the GP tree of the i-th individual in the t-th generation is $f_i^t(X)$ according to the sampling data in the k-th moving window. The fitness function of the i-th GP tree is defined as follows:

$$fitness_i^{(k)} = (x(k \times N_1 + N_2) - f_i^t(x(k \times N_1 + 1), \cdots, x(k \times N_1 + N_2 - 1))^2 \tag{9}$$

Here we are considering an existing relationship among the time sequence data. If its root mean square error(RMSE) is very small between the estimate of some individual in the t-th generation and the real data, the non-linear properties existing in the time sequence data can be expressed by this GP tree. Sort the fitness values of all individuals. When the terminal condition is reached and the first r fitness values which are the smallest are selected to represent the fitness value of the k-th moving window, that is,

$$\overline{fitness}_k = \frac{1}{r} \sum_{i=1}^{r} fitness_i^{(k)} \tag{10}$$

Comparing the fitness value $\overline{fitness}_{k-1}$ in the k-1-th moving window with the $\overline{fitness}_k$ in the k-th moving window, the improvement degree will be computed as: $F_k = \overline{fitness}_k / \overline{fitness}_{k-1}$.

When the system has the huge structural change, the improvement degree F_k will violently wave. With the acquirement of new knowledge to the system, the wave of F_k is becoming stable. Thus, the big change points will correspond to the structural change points of the system according to Artificial Intelligence. According to the statistical test or from the diagrammatic curse, we can determinate the structural change points of the complex system.

4. The Learning Algorithm based on RBF

The RBF network is a feed-forward neural network that accomplishes an input-output nonlinear mapping by a linear combination of nonlinearly transformed inputs. A normalized form of RBF network is shown in expression (11).

In expression (11),

① $X \in R^n$ is the input vector,
② $\omega_i, i = 1, \cdots, m$ are the weights,
③ $\Phi(\cdot): R^+ \to R$ are the radial basis functions (RBF),
④ $C_i \in R^n$ are the centers and Σ is a $n \times n$ matrix of width, and
⑤ m is the number of RBFs.

Its basic structure with one output node is shown in Fig. 3.

The typical RBF is the Gaussian Function that has the following form:

$$\Phi(X, C, r) = \exp(\frac{\|X - C\|^2}{2r^2}) \tag{12}$$

where its center is at C and the width is r. To simplify the problem, we often fix the width, for

$$f_m(X)=\sum_{i=1}^{m}\omega_i\Phi[(X-C_i)^T\Sigma^{-1}(X-C_i)]\Big/\sum_{i=1}^{m}\Phi[(X-C_i)^T\Sigma^{-1}(X-C_i)] \quad (11)$$

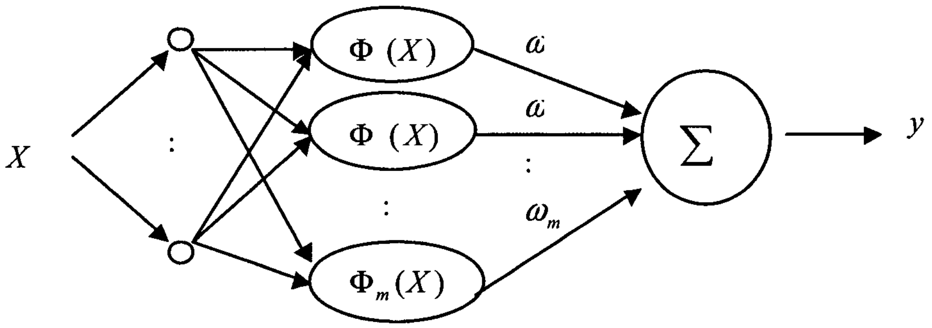

Fig.3. The basic structure of RBF model with one output node

example, let $\Sigma = I$ or $\Sigma = r^2 I$. Therefore, the determination of RBF structure depends on the determination of the centers and the linear combining weights.

The theorem about the functional equivalence of the TSK fuzzy model expressed by equation (2) and the RBF model by equation (11) has been given in reference [3]. Based on this theorem, we have developed a learning algorithm by RBF to realize the fuzzy rule extraction. RBF model has two parts to be modeled: one is the hidden layer(the antecedent part in TSK model) and the other is the output layer (the consequent part). The RBF is generally defined by its center and width, and the output part is trained by Least Squared Learning method. We have proposed an algorithm of selecting a subset of basis function from a large of candidates using stepwise regression. The main idea of the algorithm is as follows: first, the heuristic criterion called "Innovation-Contribution" is developed, and then based on this criterion and the other information criterion, we append not only the large-contributed RBF into the model but also decide whether some selected RBF should be deleted. The use of the appendable and deletable stepwise-regression algorithm makes the model structure independable of the order of RBF selected into the model which makes it possible to determine the optimal structure of RBF model. For the detail procedure of the algorithm, please refer the steps in reference [3].

The learning algorithm based on RBF is applied for every sub-stage and we can suppose that the desired output of each sub-stage is just as same as the output variable. If the network structure of the first sub-stage is determined, its output with the sampling inputs in the second sub-stage will be used to determine the model structure of the second structure, and then continue this procedure until the network structures of all the sub-stage parts are determined. By RBF network, the consequent parts of the TSK fuzzy model are the constants (corresponding to the weights $\omega_i, i=1,\cdots,m$ of RBF model). To obtain the parameters $b_j^i, j=0,1,\cdots,n$ in expression (1), LSE algorithm can be applied to find the estimates $\hat{b}_j^i, j=0,1,\cdots,n$.

5. Simulation Result

In order to evaluate the efficiency of the proposed method, we have performed the experiment on the time sequence data of synthetic indexes of closing price in Shanghai Stock Market(SHSM). There are total 223 closing price instances from Jan. 1. 1998 to Dec. 31.1998. Because of the wave data, it is necessary to have some pretreatment. Assuming that $x_t, t = 1, \cdots, N$ are the sampling data. Then compute $x_t'' = x_t' - x_{t-1}' = \log(x_t) - \log(x_{t-1})$. It approximates the relative gains of synthetic indexes in SHSM.

Simulation has been conducted to determine the structural change points by the combination of GP and the moving window, and the parameters are defined as followings: the total sampling data is 223, the delay of the time sequence is 10; the functional set of GP is (sin, cos, exp, rlog, +, -, ×, ÷); the depth of GP tree is 26; The probability of selecting leaf nodes is 0.4; the size of population and the biggest generation are 100 and 10, respectively; The cross, mutation and reproduction probability are defined as 0.9, 0.8 and 0.1. The length of representing individuals is $r = 1$.

The diagrammatic curse of the structural change parts by the MATLAB calculation is given in Fig. 4, where the cross axle shows the number of the moving windows and the ordinate expresses the improvement degrees F_k of the GP populations. Here, set the width of the moving window $N_1 = 20$, every moving step $N_2 = 3$. It indicates that each unit in the cross axle equals to a moving window of the width $N_1 = 20$. F-statistical test is used to test the structural change points of the system. When the significant level is 0.01, the critical value of F-test $F(20,20) = 1.79$. We can know from the diagrammatic curve that two huge structural change points appear at the 50-th sampling data and 152-th sampling data, respectively, because their critical value surpass F-test value. Therefore, the whole system is composed of three sub-stage: the first sub-stage is from 1 to 49 sampling data; the second is the internal [50,151] and the last belongs to the third sub-stage. Based on the RBF learning algorithm introduced by the section four, we can determine the TSK fuzzy model of the sub-stage sequentially. To illustrate the performance, we compare the IF-THEN rule number of the final model, the number of the estimated parameters and the Root Mean Square Error(RMSE) between the real data and the simulation data of the single-stage TSK fuzzy modeling method by RBF in reference [3] with those of our multi-stage method. Table 1 gives the results of the final computation.

Table 1 The comparison results of the two methods

models		indexes		
		Rules No.	Para. No.	RMSE
The single-stage model by RBF		21	126	3.641
The Multi-Stage model	1	2	12	2.4362
	2	8	65	
	3	7	57	

It seems that the proposed multi-stage model has the better performance than the single-stage model, however, on the hand, the multi-stage model reduces the problem dimension through breaking down the sampling data into multi-stages; on the other hand, as the number of the estimated parameters increases, the cost of estimated parameters is increased correspondingly. In general, multi-stage structures of the complex system can effectively reduce the problem complexity.

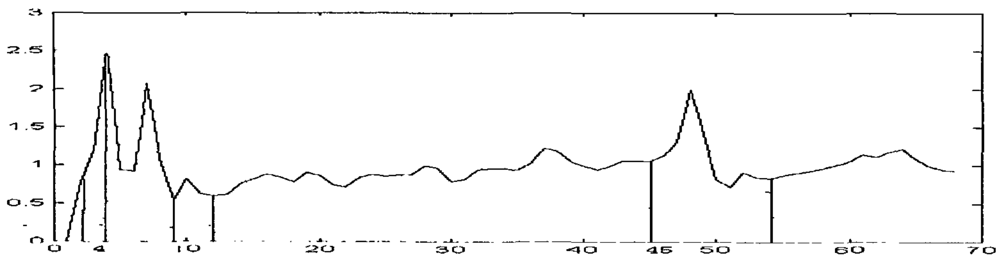

Fig. 4. The structural change curve of synthetic indexes in SHSM by GP and the moving window

6. Conclusion

To explore the dimensional problem in fuzzy complex systems, a multi-stage TSK model is proposed in this paper. This method is able to find the structural change points of input space by the combination of GP and the moving window and makes much great improvement in case of the highly dimensional non-linear system than the single stage modeling method.

Acknowledgement

This work was supported by the Brain Korea (BK) 21 project supported by the Ministry of Education of Korea.

References

1. Lin, C.T. and Lee, C. S. G, Neural-network-based fuzzy logic control and decision system, IEEE Trans. on Computer, 1991, 40(12):1320-1336.
2. Jang, J.-S., ANFIS: Adaptive-network-based fuzzy inference system, IEEE Trans. On System, Man and Cybernetics, 1993, 23:973-980.
3. Li, B. and Zhang, S., Complex Systems Modeling via Neuro-Fuzzy Method, Information and Control (in Chinese), 2001, 30 (3) : 231-233.
4. Raju, G V. S., Zhou, J., and Kisner, R. A., Hierarchical fuzzy control, International J. Control, 1991, 54:1201-1216.
5. Raju, G V. S., and Kang, GT., Fuzzy modeling and control of multi-layer incinerator, Fuzzy Sets and Systems, 1996, 18:329-246.
6. Takagi, T., and Sugeno, M., Fuzzy identification of systems and its applications, IEEE Trans. System, Man and Cybernetics, 1985, 15(1):116-132.
7. Chen, J. Q., Xi, Y. G., and Zhang, Z. J., A clustering algorithm for fuzzy model identification, Fuzzy Sets and Systems, 1998, 98:319-329.
8. Wang, L. X., Universal approximation by hierarchical fuzzy systems, Fuzzy Sets and Systems, 1998, 93:223-230.
9. Koza, J., Genetic programming III: Darwinian innovation and problem solving, Morgan Kaufman, San Francisco, CA, 1999.