한국퍼지 및 지능 시스템학회 논문지 1997, Vol. 7, No. 4, 82~92

Fuzzy Measure-based Subset Interactive Models for Interactive Systems

퍼지 측도를 이용한 상호 작용 시스템의 모델

Soon H. Kwon and M. Sugeno*
권순학, 수게노 미치오*

School of Electrical & Electronic Eng.
Yeungnam University
**Department of Computational Intelligence and Systems Science*
Tokyo Institute of Technology

요 약

본 논문에서는, 퍼지 측도와 퍼지 적분을 이용한 상호 작용 시스템의 모델 및 이의 식별법을 제시한다. 모델 식별은 다음과 같은 세 단계를 거쳐 이루어 지는데, 그 첫번째는 모델의 구조 식별이고 두번째는 식별된 구조를 갖는 모델의 파라메터 식별이다. 그리고 마지막으로는 식별된 구조와 파라메터를 갖는 모델의 최적성을 판단하여, 최적의 모델을 선정하게 된다. 본 논문에서는 최적 모델의 식별을 위하여 유전자 알고리즘 및 통계적 모델 선택 기준을 이용하여, 최적 모델들의 후보군으로부터 최적 모델을 선정하는 알고리즘을 제시한다. 본 논문에서 제시된 모델 및 이의 식별법의 타당성을 보이기 위하여, 주관적 평가 데이타 및 시계열 데이타에 적용하여 그 결과를 나타내었으며, 또한 기존의 다른 모델들로부터 얻어진 결과와 비교 검토하였다.

ABSTRACT

In this paper, a fuzzy measure and integral-based model for interactive systems is proposed. The processes of model identification consists of the following three steps : (i) structure identification (ii) parameter identification and (iii) selection of an optimal model. An algorithm for the model structure identification using the well-known genetic algorithm (GA) with a modified selection operator is proposed. A method for the identification of parameters corresponding to fuzzy measures is presented. A statistical model selection criterion is used for the selection of an optimal model among the candidates. Finally, experimental results obtained by applying the proposed model to the subjective evaluation data set and the well-known time series data are presented to show the validity of the proposed model.

1. Introduction

In recent years, we frequently encounter the term "system" not only in our daily life, but also in areas of engineering, science, business, education, and etc. One of many reasons for the wide use of the term may be that there are in great demands today to analyze concepts of systems and apply them to various situations. The word "system" is not generally used by itself, but is accompanied by an adjective or other modifier (e.g., intelligent system, interactive system, social system, etc.) which confines properties of the system to those of the adjective or modifier. For an example, the interactive system which is mainly concerned in this paper is a system that all things involved or consisted in the system act on one another.

For the analysis or prediction of characteristics of the concerned system, it is necessary to model the system on the basis of the physical law or the data obtained from the system. In case of identification of models using the given data, our attention will be given almost exclusively to choose a suitably parameterized model from the assumed class of models. The methodology of system identification involves the following three steps :

※ This research was supported by a grant from Yeungnam University

(i) selection of a class of models from which a model to represent the system is to be chosen

(ii) parameter estimation of the chosen model on the basis of the observed data

(iii) model variation based on adequate performance indices to the problem

In this paper, we will also follow these three steps.

During the last several decades, for the analysis of data with interaction, a number of linear or nonlinear models have been devised and are going on [1~6, 19]. In these models, an adjective or other modifier (e.g., linear regression model in regression analysis, threshold model in the analysis of time series data) represents the structure of the model. In spite of these research efforts, there remains the problem to represent and analyze interactions among attributes. For this reason, fuzzy measure-based models may provide reasonable and effective alternatives to classical models. Fuzzy measures proposed by Sugeno[7] are non-additive measures that provide attractive means to represent interactions among attributes. Because of their characteristics, fuzzy measures have been applied to modeling of a variety of systems [4~6].

GA's developed by Holland[8] are general purpose search procedures based on models of evolutionary processes in nature. They use operations such as selection, recombination and mutation to guide itself through the paths in the search space. Because of their robustness and the ease with which they can handle arbitrary kinds of constraints and objectives, GA's have been successfully applied to various optimization problems such as control problems, scheduling and learning systems[9].

In this paper, we propose a subset interactive model using non-monotonic fuzzy measures and the Choquet integral. GA's are used for searching a parsimonious model from a large number of possible alternative models. Two set of experimental data, one of which is the data set for nonlinear regression and the other is the well-known time series data (i.e., the Canadian lynx data), are examined to check the effectiveness of the proposed model. Experimental results show that the proposed models have superior performances compared to those using linear models and some nonlinear models.

2. Fuzzy Measures and the Choquet Integral

Fuzzy measures, including non-monotonic fuzzy measures[10], are non-additive measures and more general than the conventional Lebesgue measures assuming additivity. Hence, fuzzy measures may be appropriate for approximating processes with interactions among their inputs. We start with the definition of a fuzzy measure, which is a monotone set function. Let X be a non-empty set and let F be a σ-algebra defined on X.

Definition 1. A fuzzy measure on a measurable space (X, F) is a real-valued set function $\lambda : X \rightarrow R^+$ with the following two properties [7] :

(a) $\lambda(\phi) = 0$,

(b) $A, B \in F$ and $A \subset B \rightarrow \lambda(A) \leq \lambda(B)$,

where $R^+ = [0, \infty]$ is the set of nonnegative real numbers.

The triplet (X, F, λ) is called a fuzzy measure space and λ defined on F is called a fuzzy measure on a measurable space (X, F). The fuzzy measure can be considered as an extension of the classical probability measure (i.e., the additivity of the classical measure is replaced by the weaker condition of monotonicity). In this monotonic fuzzy measure, the fuzzy measure $\lambda(A)$ of a subset A of the universe of discourse X expresses the degree of belief/likelihood/confidence of '$x_0 \in A$', where x_0 is an unknown element of X.

However, in the real world, there exist some cases where the monotonicity is inessential. We consider some measures μ which represent degrees of interaction among subsets. Suppose A and B are disjoint subsets. If the interaction between A and B is cooperative, then $\mu(A) \leq \mu(A \cup B)$ and/or $\mu(B) \leq \mu(A \cup B)$ will be satisfied. But if the interaction between A and B is countervailing, then $\mu(A) \geq \mu(A \cup B)$ and/or $\mu(B) \geq \mu(A \cup B)$ may be satisfied, which violate the monotonicity condition of the fuzzy measure. Thus, it is natural that we consider a non-monotonic fuzzy measure[10].

Definition 2. A non-monotonic fuzzy measure μ on a measurable space (X, F) is a real-valued set function $\mu : X \rightarrow R$ satisfying $\mu(\phi) = 0$.

Throughout this paper, we use the interpretation that the non-additivity of the fuzzy measure expresses the interaction among subsets. We can define the Choquet integral with respect to non-monotonic fuzzy measures[10], which is very reasonable as an integration with respect to fuzzy measures.

Definition 3. The Choquet integral, denoted by the symbol, $(c)\int$, of a measurable function $f : X \to R$ with respect to a non-monotonic fuzzy measure μ over a set $A \in F$ is defined by

$$(c)\int_A f d\mu \equiv \int_0^{\infty} \mu(\{x \mid f(x) \geq r\} \cap A)dr + \int_{-\infty}^{0} [\mu(\{x \mid f(x) \geq r\} \cap A) - \mu(A)]dr \quad (1)$$

where the integral in the right side is an ordinary Lebesgue integral. A measurable function f is said to be integrable iff the Choquet integral of f over X is finite. For the special case of μ being additive, the definition above coincides with the Lebesgue integral. The following are basic properties of the Choquet integral with respect to a non-monotonic fuzzy measure μ.

$$(c)\int 1_A \, d\mu = \mu(A) \quad (2)$$

where, $1_A = \begin{cases} 1 & x \in A, \\ 0 & otherwise \end{cases}$

$$(c)\int(af+b)d\mu = a \cdot (c)\int f d\mu + b \cdot \mu(X) \; \forall a,b \in R \,. \quad (3)$$

3. Subset Interactive Models

As pointed out in section 2, the non-monotonic fuzzy measure is appropriate for modeling a process with interactions which are cooperative and/or countervailing. In this section, we propose a new approach to modeling of interactive systems using non-monotonic fuzzy measures and the Choquet integral, and present a subset interactive model which may be a generalized model of the classical linear regression or time series models.

If we write down a general nonlinear subset interactive model of order n, it takes the following form :

$$y = h(x_1, x_2, \ldots, x_n) + e, \quad (4)$$

where y is an output, h is a nonlinear function of input variables (i.e., values of attributes $x_1, x_2, \ldots, x_n$), and e is the white noise series (i.e., a sequence of independent zero mean and finite variance random variables). If we assume that the structure of regression model can be described by a linear form, the model (4) will be a linear regression model written as

$$y = \sum_{i=1}^{n} a_i x_t + e_t \,, \quad (5)$$

and if we assume that the structure of the time series can be described by a linear form, the model (4) will be a linear AR(Auto-Regressive) model written as

$$x_t = \sum_{i=1}^{n} a_i x_{t-i} + e_t \,. \quad (6)$$

However, the assumption of linearity in this model is a very strict one. As Tong [3] pointed out, a linear model is totally inadequate as a tool to analyze more intricate phenomena which are apparent in subjective evaluation data or in time series data. Therefore, it is very natural to remove the assumption that inputs are independent of each other, and then apply fuzzy measures to express the interaction among inputs. In this case, it is necessary to devise methods which overcome difficulties caused by high dimensionality (i. e., the curse of dimensionality that the data points are sparse in the high dimensional sample space and the difficulty of interpretation). In another respect, subset models are often desirable, especially when the data exhibits some form of periodic behavior. In such cases, fitting full order models often results in the fitted coefficients of some lags being close to zero.

With this background, we will propose a subset interactive model which is a natural extension of linear regression models or linear AR time series models. To give a more precise definition of the model, let $X = \{x_1, \ldots, x_n\}$ be a set of explanatory variables, let f be a real-valued function of X, and let x_k be a response variable. The interactions among explanatory variables can not always be described by additive or monotonic functions. In such cases, the fitting ability of a model using additive or monotonic functions will be degraded. With this background and notation, we will propose a new model for the interactive system as follows:

$$y = \sum_{i=1}^{p} (c)\int_{C_i} f \, d\mu + e \,, \quad (7)$$

where (i) C_i is an element of a class $C = \{C_1, \ldots, C_p\}$ such that , $C_i \subset S$, $C_i \neq \phi$ and $\bigcup_{i=1}^{p} = C_i = S$ for i = 1,... , p and $S \subset X$, $C \neq \phi$, (ii) m is a non-monotonic fuzzy measure, (iii) f is a measurable function $f: X \rightarrow R$, and (iv) errors e are assumed to be independent and normal, with mean zero and variance σ^2.

The cognitive interpretation of this model is that the classification of explanatory variables X can be accomplished with the strategy that the explanatory variables included in the same class have strong interactions, but the relationships among classes may be characterized by weak interactions. Fig. 1 shows a structure of the proposed model.

The proposed model coincides with the linear regression or AR model if the non-monotonic fuzzy measures are additive, and it becomes a subset AR model when the non-monotonic fuzzy measures are additive and the subset S is not equal to X. Thus, to identify the subset interactive model (7) from the given data, we have to estimate non-monotonic fuzzy measures m and a covering C (i.e., a structure of the model).

We will consider how to identify non-monotonic fuzzy measures for the developed model. In this paper, the parameters (i.e., non-monotonic fuzzy measures μ) are identified by using the maximum-likelihood method which yields the estimate μ that maximizes the log-likelihood function $L(\mu, \sigma^2)$ given by

$$L(\mu,\sigma^2) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(Z_i - \sum_{j=1}^{P}(c)\int_{C_j} fd\mu\right)^2. \quad (8)$$

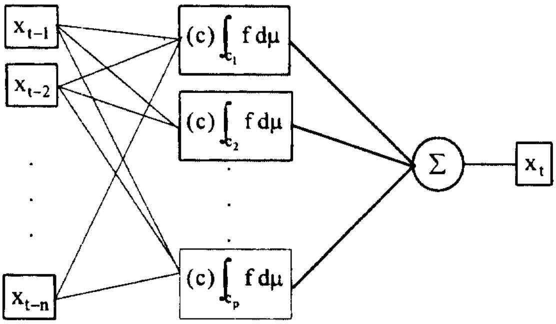

Fig. 1. A structure of the proposed model.

where N is the number of data. Thus maximum-likelihood estimates μ will be obtained by minimizing a variance (9), which can be solved by use of the least squares method.

$$L(\mu,\sigma^2) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Z_i - \sum_{j=1}^{P}(c)\int_{C_j} fd\mu\right)^2. \quad (9)$$

4. Subset Selection using Genetic Algorithms

The structure identification of (7) can be considered a combinatorial optimization problem (i.e., to search a covering C of X). In order to obtain good (sub optimal) solutions for such combinatorial problems at a low computational cost, many algorithms have been devised and used [1]. One of the most comprehensive, but cumbersome ways to solve it is to examine all possible combinations of subsets of X (i.e., exhaustive examination). However, the straightforward search for structure and non-monotonic fuzzy measures optimizing a cost function over all possible coverings may be a computationally hard problem when the number of inputs increases.

Genetic algorithms which evolve according to rules of evolving operators perform multi-directional searches by maintaining a population of potential solutions. They are fundamentally different from conventional search methods in facts that they perform stochastic, multi-directional, and parallel search in the search space. They use operations such as selection, recombination and mutation to guide itself through the paths in the search space.

The simple GA (SGA) [9] has a serious problem called premature convergence, where a single population element dominates the population and the capability of the system to explore the search space is impaired. In the following, we will briefly explain each of the GA components and present some strategies which allow the use of a relatively small population size and prevent premature convergence.

Potential solutions (i.e., coverings of X) are encoded into binary based-strings of which j-th bit (i.e., gene) represents whether the j-th subset of X is included in the covering or not. The j is obtained by

$$j = \sum_{x_i \in A,\ A \in C} 2^{i-1}. \tag{10}$$

For example, the string representing a covering C = {{x_1, x_2}, {x_1, x_3}} of X = {x_1, x_2, x_3} is as follows :

subsets {x_1} {x_2} {x_1,x_2} {x_3} {x_1,x_3} {x_2,x_3} {x_1,x_2,x_3}
string (1 1 1 1 1 0 0).

For convenience, we denote the string of a covering C as *bit* (C). For example, *bit* (C_1 = {{x_1}}) represents the string (1 0 0 0 0 0 0) , and *bit* (C_2 = {X}) represents the string (1 1 1 1 1 1 1). In this n bit string representation, there are 2^n possible combinations. However, some of these are infeasible (e.g., B = (1 0 1 0 0 0 0)). Thus, it is necessary to map these infeasible strings into other feasible strings by using lower and upper approximation as the mapping. The lower approximation of a string B_* is a string which represents a maximal irreducible covering that belongs to B. The upper approximation of a string B^* is a string which represents a minimal irreducible covering that overlaps with B. For example, we consider the above bit string B = (1 0 1 0 0 0 0). Then, the lower approximation of B is B_* = (1 0 0 0 0 0 0), and the upper approximation of B is B_* = (1 1 1 0 0 0 0). These lower and upper approximations of a string make all possible strings be feasible, though these may lead to probabilistic selection bias. The probabilistic selection bias problem can be overcome by a selection strategy to be discussed below.

The fitness function plays the role of the environment, rating potential solutions in terms of their fitness. Based on the characteristics of the problem to be solved, we use Bayesian Information Criterion (BIC) [11] given by

$$\text{BIC} = N\log \hat{\sigma}^2 + m \log N, \tag{11}$$

where N is the number of data, m is the number of independent parameters and log denotes natural logarithm. It is well known fact that BIC provides a consistent model selection. Thus, the use of BIC as a model selection criterion prevents over-parameterization that may be occurred in the use of inconsistent model selection criteria.

The primary work of the genetic algorithm is performed by three operators, select, crossover, and mutation. Such genetic operators, which are applied to partial or entire populations at each generation, are designed to be types of problem specific operators. Any genetic operator should pass some chromosome structures from parent to offspring and preserve population diversity.

For the simple GA, the competition is not between parents and their offspring but only between the entire set of offspring. And the probability-based selection of the SGA has a possibility that the chromosome with the highest fitness does not survive to the next generation. These may lead to the loss of population diversity, so that the solution falls into a local optimum. To combat this phenomenon, we devise a selection scheme which is capable of forming and maintaining stable subpopulations, or niches as follows. As the topology of the selection scheme, we adopt a ring structure shown in Fig. 2.

In Fig. 2, $p_i(t)$ (i = 1, 2, ... , n) represent chromosomes at time t, P(t) = {$p_1(t)$, $p_2(t)$, ..., $p_n(t)$} represents a population at time t, and each niche has only one chromosome. Thus, if the size of the initial population is n, the number of niches becomes n. The selection operator for mating performs simple deterministic selection without replacement from the population $p_i(t)$ (i.e., if a single individual $p_i(t)$ selected, then the other individual selected is $p_{i+1}(t)$ (also, the neighbor of $p_n(t)$ is $p_1(t)$). This selection method has a capability that makes probabilistic selection bias negligible due to the representation scheme mentioned above. If $p_i(t)$ and $p_{i+1}(t)$ are selected as parents, then the best among offspring (i.e., $c_i(t)^*$, $c_i(t)^*$, $c_{i+1}(t)^*$ and

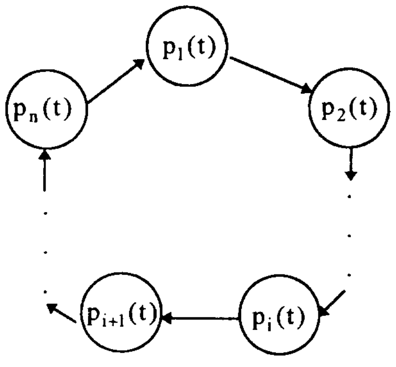

Fig. 2. Ring topology for niching mechanism.

$c_{i+1}(t)^*$) and $p_i(t)$ replaces the $p_i(t)$ in the next generation. That is, the best solution is copied into the next generation, called elitist selection.

The multiple points crossover operation is performed in the following way. First, chromosomes in the pool are mated according to some rules, then each pair of mated chromosomes crosses over genetic information from one chromosome to another. The crossover probability of 0.6 is fixed for all trials.

Mutation arbitrarily alters one or more genes of a selected chromosome by a random change with a probability equal to the mutation probability, which introduce some innovations into the population. The swap type mutation scheme is used with a mutation probability of 0.1, fixed for all trials.

5. Applications

To examine the adequacy of the subset interactive model using non-monotonic fuzzy measures and the Choquet integral, we apply it to some real data (i.e., subjective evaluation data : sensory evaluation data of rice taste and the well-known time series data : the Canadian lynx data) and show experimental results obtained from these data. Simultaneously, we apply traditional models to these data set, and compare fitting and forecasting performances of these models.

5.1. Subjective evaluation data

The design strategy of designers or experts faced with developments of new products depends greatly on not only its physical properties but also consumer preference. We may find typical examples such as rice, blended coffee, cars, cosmetics and etc. around our daily life. As consumer preference is based on the subjective evaluations on something, it is important to investigate how the subjective evaluation is done. As the first application of our model, we consider the sensory evaluation of rice taste to which people for whom rice is a staple are very sensitive.

Let (X,Y) be a pair of given data such that $X = \{x_1, \ldots, x_n\}$ is a set of attributes to be considered, and the overall evaluation Y is real-valued. To show characteristics of the proposed model more quantitatively, we compare it with a linear regression model given by

$$y = \sum_{i=1} (c) \int_{C_i} f d\mu + e \tag{12}$$

where $C_1 = \{x_1\}$, $C_2 = \{x_2\}$, . . . , $C_n = \{x_n\}$ and a parametric non-structurized subjective evaluation model, in short PSE-model, using non-monotonic fuzzy measures and the Choquet integral given by

$$y = \sum_{i=1}^{1} (c) \int_{C_i} f d\mu + e \tag{13}$$

where the covering $C_1 = \{x_1, \ldots, x_n\}$.

In this application, we used 104 data points given by sensory evaluaton of 24 panelists with excellent evaluation ability. The data consists of an overall evaluation (Y), and 5 additional partial evaluations which are believed to be related to the taste of rice, that is, x_1 : flavor, x_2 : appearance, x_3 : taste, x4 : stickiness and x5 : toughness. Sensory evaluation was done with 11 step scales for flavor, appearance, taste and overall evaluation. On the other hands, 7 step scales are adopted for stickiness and toughness. All data were preprocessed to make all variables have values within the range of 0.0 to 1.0 and have 0.5 as the average of each variable. The data set was divided into two halves, called odd-half and even-half. The odd-half was used as the training data for identification of the model, and the even-half was used as checking data.

A subset interactive model shown in (14) was identified using the GA-based method with BIC as a performance index.

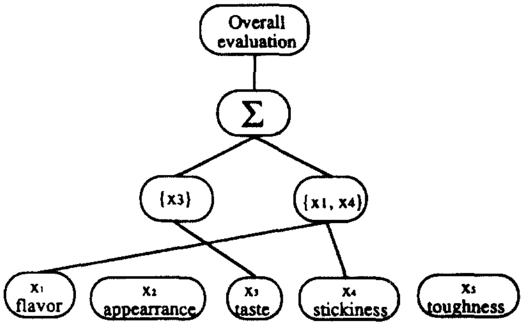

Fig. 3. A schematic diagram of the identified model.

Table 1. Identified non-monotonic fuzzy measures

subset A	fuzzy measures μ(A)
ϕ	0.0000
$\{x_1\}$	0.0552
$\{x_3\}$	0.5164
$\{x_4\}$	0.2577
$\{x_1, x_4\}$	0.5132

Table 2. Performance table

models	$\hat{\sigma}_t$	number of parameters	BIC	$\hat{\sigma}_c$
linear	0.00078	6	-348.38	0.00174
PSE	0.00028	32	-299.93	0.00720
subset interactive	0.00056	5	-369.21	0.00154

$$y=\sum_{i=1}^{2}(c)\int_{C_i} f d\mu+e \qquad (14)$$

where C1 = {x3}, C2 = {x1, x4}. Fig. 3 shows a schematic diagram of the identified model. Table 1 shows identified non-monotonic fuzzy measures.

From the model (14), we can see that the overall evaluation Y may be influenced by three attributes among five attributes, that is, x_1 : flavor, x_3 : taste, and x_4 : stickiness. Table 2 shows experimental results obtained by applying the training data to a linear model, a PSE-model and a subset interactive model, respectively. The second and third columns in table 2 give mean sums of squared errors obtained from the training data and the number of parameters for each model. In column four, values of BIC are listed. The last column shows values of mean sums of squared errors obtained by applying identified models to the checking data.

From table 2, we can see that the proposed subset interactive model provides the best compromise between the goodness of fit and the complexity of the model. Fig. 4 shows the relationship between the estimated and the observed overall evaluations obtained from the training data (a), and the checking data (b).

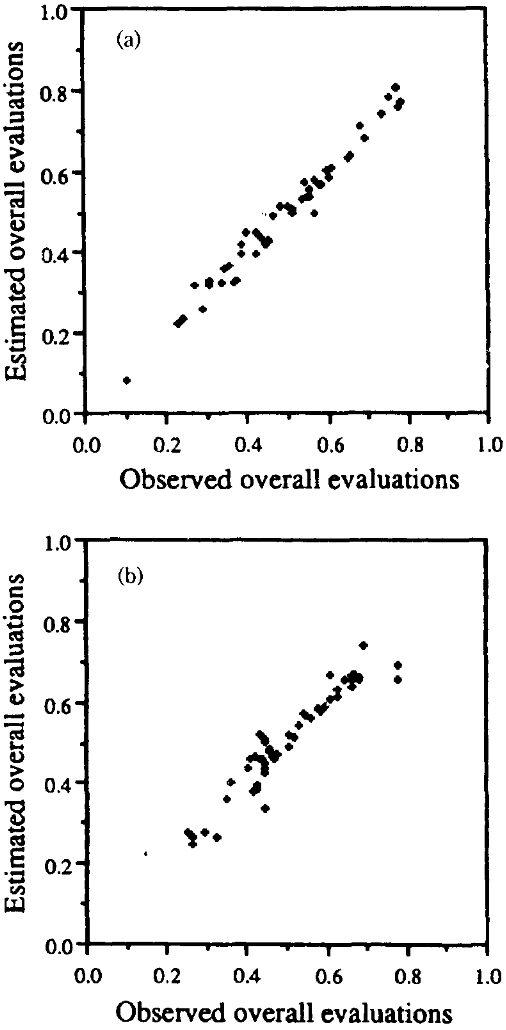

Fig. 4. Relationship between the observed versus the estimated overall evaluations
(a) the training data (b) the checking data

5.2. Time series data

As the second application, we consider the Canadian lynx data set. This data set, which consists of 114 observations, shows the annual record of the number of Canadian lynx trapped in the Mackenzie river district of Canada for the years 1821-1934. It has been analyzed by many time series analysts to validate their time series models since Moran [12], who initially applied a logarithmic transformation (i.e., log10) in order to reduce the asymmetrical appearance of the original data. For example, Campbell and Walker [13], Tong [14], Tong and Lim [15], Gabr and Subba Rao [16], Haggan [17] are the representational works, of which models have been shown to has the advantage of accounting for many nonlinear features, and to provide good fits and forecasting to this data.

To compare fitting and forecasting performances obtained from several models with the subset interactive model, we partition the logarithmically transformed data into a data set with the first 100 observations (observations over the period 1821-1920) for fitting and the other data set with the last 14 observations (observations over the period 1821-1934) for prediction. We will compare the relative fit of linear and nonlinear models using performance indices such as the mean sum of squares of residuals and the normalized AIC [18] values given by

$$NAIC = \log\hat{\sigma}^2 + \frac{2m}{N}, \tag{15}$$

where N is the number of data, m is the number of independent parameters and log denotes natural logarithm. It is well-known fact that AIC is a criterion to measure the quality of a model by its goodness of fit for given data and by its complexity. That is, a model selected by minimizing the value of AIC has the best compromise between goodness of fit for given data and model complexity.

Since the best AR model is found to be AR(12), the maximum lag is chosen to be 12 (the first 12 observations are omitted). The fitted full AR model, the best subset AR model, the best subset bilinear model, the threshold subset interactive model and the subset interactive model are as follows:

1) Full AR model [16]

The fitted model to the mean corrected observations is

$$\begin{aligned} x_t &- 1.0541x_{t-1} + 0.4539x_{t-2} - 0.32597x_{t-3} + 0.37912x_{t-4} \\ &- 0.23452x_{t-5} + 0.17570x_{t-6} - 0.09598x_{t-7} \\ &+ 0.128437x_{t-8} - 0.27435x_{t-9} - 0.11427x_{t-10} \\ &+ 0.18534x_{t-11} + 0.17128x_{t-12} = e_t. \end{aligned} \tag{16}$$

The mean sum of squares of residuals is 0.0358 and the AIC value is -266.9 (the normalized AIC is -3.033).

2) Best subset AR model [16]

The fitted best subset AR model to the mean corrected observations is

$$\begin{aligned} x_t &- 1.01705x_{t-1} + 0.39997x_{t-2} - 0.25851x_{t-3} \\ &+ 0.22037x_{t-4} - 0.21099xt\text{-}9 + 0.25343x_{t-12} = e_t. \end{aligned} \tag{17}$$

The mean sum of squares of residuals is 0.0378 and the AIC value is -274.2 (the normalized AIC for subset AR model is -3.116).

3) Best subset bilinear model [16]

The fitted best subset bilinear model is

$$\begin{aligned} X_t &- 0.77227X_{t-1} + 0.091572X_{t-2} - 0.083073x_{t-3} \\ &+ 0.261493x_{t-4} - 0.225585X_{t-9} + 0.245841x_{t-12} \\ &- 1.486292 = - 0.7893X_{t-3}\, e_{t-9} + 0.4798Xt\text{-}9e_{t-9} \\ &+ 0.3902X_{t-6}e_{t-2} + 0.1326X_{t-1}e_{t-1} + 0.07944X_{t-2}e_{t-7} \\ &- 0.3212X_{t-4}e_{t-2} + e_t. \end{aligned} \tag{18}$$

The mean sum of squares of residuals is 0.0223 and the AIC value is -308.7 (the normalized AIC for bilinear model is -3.508).

4) The threshold autoregressive model [15]

The self-exciting threshold autoregressive model SETAR (2;6;3) fitted by Tong and Lim is

$$Xt = \begin{cases} 0.8023 + 1.0676X_{t-1} - 0.2069X_{t-2} + 0.1712X_{t-3} - 0.4528X_{t-4} + 0.2237X_{t-5} - 0.331X_{t-6} + e_t & \text{if } 0 < X_{t-2} \le 3.50, \\ 2.2964 + 1.4246X_{t-1} - 1.0795X_{t-2} - 0.0907X_{t-3} + e_t & \text{if } 3.05 < X_{t-2} \le 10 \\ 0 & otherwise \end{cases} \tag{19}$$

The one-step-ahead predictors of this model are given in their paper. The mean sum of squares of residuals is 0.0415 and the AIC value is -262.7 (the normalized AIC for SETAR (2;6;3) is -2.985).

5) Subset interactive autoregressive model

For reducing the computational cost, we choose 9 variables (i.e., $X = \{x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-9}, x_{t-10}, x_{t-11}, x_{t-12}\}$) among the given 12 variables (i.e., $\{x_{t-1}, x_{t-2}, x_{t-3}, \ldots, x_{t-12}\}$). The bit string length of 511 is determined by the number of subsets of X. The population size is 500 and the number of subpopulations is 500. The crossover probability is 0.6 and the mutation probability is 0.1. The fitted subset interactive autoregressive model to the mean corrected observations is

$$y = \sum_{i=1}^{4} (c) \int_{C} f d\mu + e \tag{20}$$

where the identified covering C is $\{C_1 = \{x_{t-1}\}, C_2 = \{x_{t-2}, x_{t-10}\}, C_3 = \{x_{t-4}, x_{t-10}\}, C_4 = \{x_{t-12}\}\}$, and the identified non-monotonic fuzzy measures m are exhibited in Table 3. As we can see in Table 3,

Table 3. Identified non-monotonic fuzzy measures

subset A	fuzzy measures $\mu(A)$
ϕ	0.0000
$\{x_{t-1}\}$	0.8961
$\{x_{t-2}\}$	-0.4333
$\{x_{t-4}\}$	0.1713
$\{x_{t-10}\}$	0.0491
$\{x_{t-2}, x_{t-10}\}$	0.0843
$\{x_{t-4}, x_{t-10}\}$	-0.2596
$\{x_{t-12}\}$	-0.2747

the number of independent parameters is 8. And the suboptimal subset is $S = \{x_{t-1}, x_{t-2}, x_{t-4}, x_{t-10}, x_{t-12}\}$. The mean sum of squares of residuals is 0.0333 and the AIC value is -283.281 (the normalized AIC for this subset interactive autoregressive model is -3.219).

The comparison graph of observations and simulated data is plotted in Fig. 5.

The performance of time series models may be judged on the basis of their forecasting performances. In order to compare the forecasting performance of the fitted models, we obtained the one-step-ahead predictions for the period up to 1955 given in Table 4.

The mean sum of squares of one-step-ahead prediction errors for the full AR model is 0.025488, for the best subset AR model is 0.022328, for the best subset bilinear model is 0.013306, and for the subset interactive autoregressive model is 0.02721. From

Table 4. One-step-ahead predictions of the Canadian

t	X(t)	Full AR	Subset AR	SBL	SI AR
1921	2.360	2.389	2.362	2.410	2.211
1922	2.601	2.812	2.792	2.745	2.731
1923	3.054	2.788	2.863	2.911	2.910
1924	3.386	3.197	3.206	3.211	3.195
1925	3.553	3.354	3.338	3.341	3.366
1926	3.468	3.431	3.303	3.438	3.488
1927	3.187	2.860	2.946	3.152	2.937
1928	2.723	2.624	2.636	2.569	2.448
1929	2.686	2.485	2.435	2.796	2.358
1930	2.821	2.853	2.832	2.825	2.183
1931	3.000	2.973	2.978	3.056	3.040
1932	3.021	3.255	3.262	3.175	3.173
1933	3.424	3.397	3.425	3.291	3.392
1934	3.531	3.563	3.562	3.444	3.524

this results, we can see that the subset interactive autoregressive model does not have good performance of prediction in comparison with other models. For this reason, we note that the model selection criterion used in this paper (i.e., BIC) may not be adequate for this purpose, so that the fitted subset interactive autoregressive model is not the best but sub optimal model.

6. Conclusions

In this paper, we proposed a subset interactive model using non-monotonic fuzzy measures and the

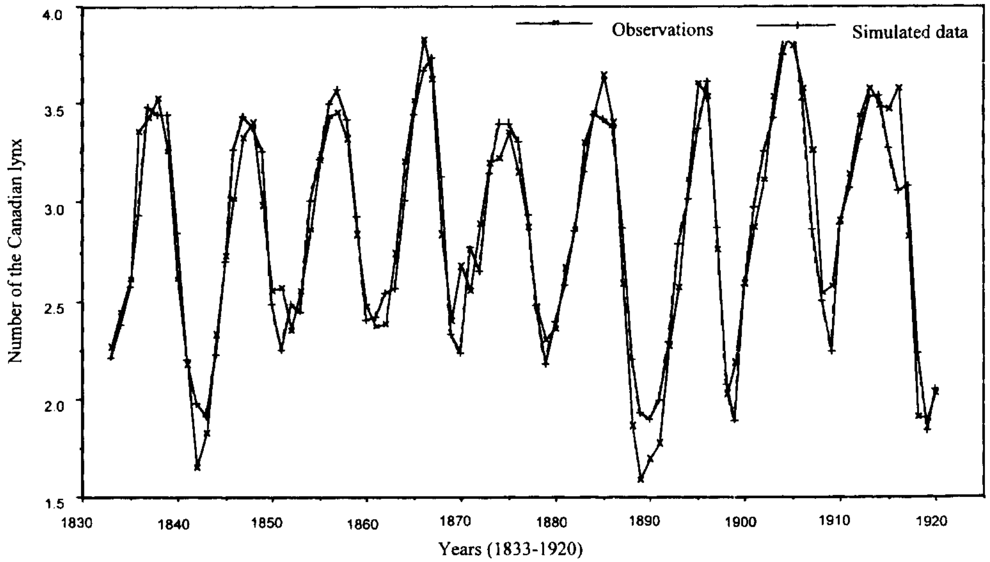

Fig. 5. Comparison of observations and simulated data.

Choquet integral. We discussed the relationship between the proposed model and traditional models such as regression models and time series models. Furthermore, we have presented a subset selection method that uses genetic algorithms. The devised subset selection algorithm has a niching mechanism for preserving the population diversity and then preventing the solution from falling into a local optimum. Finally, we have shown that the suggested model performed well in comparison with some other models (e.g., linear regression model, the full AR model).

References

[1] R. L. Mason, and R. F. Gunst, Regression Analysis and its Application, Marcel-Decker, New York, 1980.

[2] M. B. Priestley, Spectral analysis and time series, Academic Press, New York, 1981.

[3] H. Tong, Non-linear time series : a dynamical system approach, Oxford Univ. Press, Oxford, 1990.

[4] M. Sugeno and S. H. Kwon, "A clusterwise regression-type model for subjective evaluation," Japanese J. of Fuzzy Theory and Systems, vol. 7, no. 2, pp.291-310, 1995.

[5] M. Sugeno and S. H. Kwon, "A new approach to time series modelling with fuzzy measures and the Choquet integral," Proc. FUZZ-IEEE/IFES '95, Yokohama, Japan, pp. 799-804 , March 1995.

[6] K. Tanaka, and M. Sugeno, "A Study on Subjective Evaluations of Printed Color Images," Int. J. Approximate Reasoning 5 pp. 213-222, 1991.

[7] M. Sugeno, Theory of fuzzy integral and its applications, Doctoral Thesis, Tokyo Institute of Technology, 1974.

[8] J. H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.

[9] D. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, Mass., 1989.

[10] T. Murohushi, M. Sugeno, and M. Machida, "Non-monotonic fuzzy measures and the Choquet integral," Fuzzy Sets and Systems 64, pp. 73-86, 1994.

[11] G. Schwarz, "Estimating the dimension of a model," Ann. Statist. Vol. 6, pp. 461-464, 1978.

[12] P. A. P. Moran, "The statistical analysis of the Canadian lynx cycle," I. Austr. J. Zool., Vol. 1, pp. 163-173, 1953.

[13] M. J. Campbell, and A. M. Walker, "A survey of statistical work on the Mackenzie river series of annual Canadian lynx trappings for the year 1821-1934 and a new analysis," J. Roy. Statist. Soc. A 140, pp. 411-431, 1977.

[14] H. Tong, "Some comments on the Canadian lynx data," J. Roy. Statist. Soc. A 140, pp. 432- 436, 1977.

[15] H. Tong, and K. S. Lim, "Threshold autoregression, limit cycles and cyclical data," J. Roy. Statist. Soc. Ser. B 42, pp. 245- 292, 1980.

[16] M. M. Gabr, and T. Subba Rao, "The estimation and prediction of subset bilinear time series models with applications," J. Time Series Anal. Vol. 2, pp. 155-171, 1981.

[17] V. Haggan, and O. B. Oyetunji, "On the selection of subset autoregressive time series models," J. Time Series Anal. Vol. 5, pp. 103-113, 1984.

[18] H. Akaike, "A new look at the statistical model identification," IEEE Trans. on Automatic Control 19, pp. 716-723, 1974.

[19] G. E. P. Box, and G. M. Jenkins, Time series analysis, Forecasting and control, 2nd ed., Holden-Day, San Francisco, 1976.

Michio Sugeno 준회원

Michio Sugeno was born in Yokohama, Japan. He received the B.S. degree in physics from the University of Tokyo in 1962. From 1962 to 1965 he worked in nuclear engineering for Mitsubishi Atomic Power Industry.

From 1965 to 1976 he was a Research Associate in control engineering at the Tokyo Institute of Technology, Tokyo. He received the doctor of engineering degree from the Tokyo Institute of Technology in 1975. In 1976 he joined the Graduate School of the Tokyo Institute of Technology, Yokohama, as an Associate Professor. He is currently a Professor in department of computational intelligence and systems science. Since 1989 he has been a Leading Advisor at the Laboratory for International Fuzzy Engineering Research, Yokohama. Since 1991 he has been the president of the Japan Society for Fuzzy Theory and Systems. He is currently the president of the IFSA. His research interests include fuzzy computing based on systemic functional grammar and fuzzy logic, fuzzy control of helicopters, and fuzzy measure analysis.

Soon Hak Kwon 종신회원

Soon Hak Kwon was born in Taejon, Korea. He received the B.S. and the M.S. degrees in control engineering from Seoul National University in 1983 and 1985, respectively. He received the doctor of engineering degree from the Tokyo Institute of Technology in 1995.

From 1986 to 1991 he was a researcher in control engineering at Korea Institute of Science and Technology (KIST). Dr. Kwon has been with the School of Electrical and Electronic Engineering at Yeungnam University, Kyongbuk, Korea as an assistant professor since 1996. His research interests include modeling and control of intelligent systems.