Reduced-Set Vector-Based Interval Type-2 Fuzzy Neural Network

LONG YU, JIAN XIAO, SONG WANG
School of Electrical Engineering
Southwest Jiaotong University
Chengdu
CHINA
yulong.swjtu@163.com

Abstract: - This paper describes an interval type-2 fuzzy modeling framework, reduced-set vector-based interval type-2 fuzzy neural network (RV-based IT2FNN), to characterize the representation in fuzzy logic inference procedure. The model proposed introduces the concept of interval kernel to interval type-2 fuzzy membership, and provides an architecure to extract reduced-set vectors for generating interval type-2 fuzzy rules. Thus, the overall RV-based IT2FNN can be represented as series expansion of interval kernel, and it does not have to determine the number of rules in advance. By using a hybrid learning mechanism, the proposed RV-based IT2FNN can construct an input-ouput mapping from the training data in the form of fuzzy rules. At last, simulation results show that the RV-based IT2FNN obtained possesses nice generalization and transparency.

Key-Words: - interval type-2; fuzzy modeling; reduced-set; interval kernel; insensitive learning

1 Introduction

Both fuzzy logic and neural networks are aimed at exploiting human knowledge processing capability. The fuzzy logic system using linguistic information can model the qualitiative aspects of human knowledge and reasoning processes without employing precise quantitative analyzes. The neural network is a popular generation of information processing systems that equip the intelligent methodologies with better learning capabilities. Much research has been done on fuzzy neural networks, which combine the capability of fuzzy reasoning in handling uncertain information and the capability of neural networks in learning from processes [1].

By far, there are many neural learning mechanisms that has been designed to construct fuzzy model, especially several authors had attempted to incorporate support vector machine (SVM) which is known as a new neural network in statistical learning theory into fuzzy modeling [2]-[7]. To combine them, thus taking the advantages of both approaches, appears to be desirable. For example, SVM were successfully used to construct the fuzzy classifiers [2], [3]. By utilizing this methodology, support vector learning mechanisem had also been applied to identify the structure of fuzzy logic systems [4], [5]. Furthermore, considering too many support vectors for generating fuzzy rules, some other advanced methods aimed at extracting reduced-set vectors for generating rules [6], [7].

However, one point must be noticed that nearly all the approaches above about the combination of SVM and fuzzy logic systems are related to type-1 fuzzy logic. Mendel had claimed that type-1 fuzzy logic is unable to handle uncertainties which actually lie in real world, whereas the type-2 fuzzy logic can model them and minimize their effects [8]. Hence, although the advanced learning mechanisms are proved to be adequate, the nature of type-1 fuzzy logic hampers itself to deal with those complicated systems which involves more uncertain information. Therefore, this led to several type-2 fuzzy logic fuzzy modeling algorithms, especially the interval form owing to its decreasing computational intensity [9]-[15].

Liang *et al*. [9] firstly employed steepest-descent method to tune the parameters of interval type-2 fuzzy logic system (IT2FLS). After that, some researchers attempted to integrate neural network into IT2FLS in order to take advantages of both of aspects [10]-[12]. In [10] and [11], type-2 fuzzy sets were used in combination with self-learning dynamics of membership functions. In [12], a model named type-2 fuzzy neural network (T2FNN) was presented with dynamical optimal training, and the stable and optimal learning rate was derived for each iteration in the training process. Furthermore, according to the learing mechanism of the well-known model, adaptive neuro-fuzzy inference system (ANFIS), a hybrid learning algorithm for IT2FLS was proposed. In the forward pass, the consequent parameters were estimated by recursive least-squares (RLS) method. In the backward pass, the error propagates backward, and the antecedent parameters were estimated by back-propagation (BP) method [13].

Although these interval type-2 fuzzy modeling approaches are very effective in solving actural problems described by numerical examples, their use may be hampered by the selection of the initial structure of IT2FLS available to the designer, which require the choice of the option more suited to the application of interest. Actually, the usual way about the selection is a tedious and time consuming process of successive trials. Taking the number of fuzzy rules as the example, too many of them can lead to a complicated IT2FLS which will lose its generalization ability, whereas few of them can not make it possess a good approximation ability. Consequently, the IT2FLS, built on finite amount of given training data, can not generalize best if the right trade-off isn't found between accuracy and the capability of fuzzy model set. The capability, here, could be understood as the number of fuzzy rules in rule base. Hence, a good type-2 fuzzy rule base should have a small number of rules so as to make it more transparency and interpretability. To solve this problem, the work presented in this paper will attempt to introduce reduced-set vector machine (RVM) which is regarded as a simlified version of SVM into the field of interval type-2 fuzzy logic, and present a new kernel function, interval kernel function, to the interval type-2 fuzzy membership. In result, a new structure of IT2FLS with hybrid learning mechanism is formulated in an automatic procedure.

In details, the model proposed is named as RV-based IT2FNN. It combines IT2FLS with RVM by means of interval kernel, and fuses two mechanism into a new interval type-2 fuzzy inference structure. The learning of RV-based IT2FNN is realized via a hybrid learning mechanism involving two sub-algorithms: bottom-up simplification algorithm, which is exploited to extract reduced-set vectors for rapidly generating interval type-2 fuzzy IF-THEN rules and ε insensitive learning algorithm, which is employed to tune the weighted parameters of RV-based IT2FNN.

The rest of this paper is organized as follows: traditional IT2FLS and RVM are briefly summarized and discussed in Section 2. A new concept of interval kernel is proposed, and the relation between interval kernel and interval type-2 fuzzy membership is studied in Section 3. Section 4 describes the new interval type-2 fuzzy model proposed, and presents the associated learning mechanism. Some numerical results about the proposed interval type-2 fuzzy model are shown in Section 5. Finally, Section 6 concludes this paper.

2 Review of earlier works and analysis

Given crisp input and output observation data from an unknown system, data-driven methods aim to construct a decision function $F(\mathbf{x})$ that can serve as an approximation of the system. Indeed, both of IT2FLS and RVM are employed to describe the decision function. This section reviews the methodologies of IT2FLS and RVM.

2.1 IT2FLS

IT2FLS characterizes the system by a collection of interpretable interval type-2 if-then fuzzy rules, and the traditional IT2FLS consists of a set of rules with the following structure [8]:

$$R_i: \textit{If } x_1 \textit{ is } \tilde{A}_{i1} \textit{ and } x_2 \textit{ is } \tilde{A}_{i2} \textit{ and} \cdots x_d \textit{ is } \tilde{A}_{id}, \quad \textit{then} \quad y_i = [y_{il}, y_{ir}] \quad \textit{for } i = 1, 2 \cdots, c. \tag{1}$$

In (1), R_i is the ith rule in the interval type-2 rule base, parameter d is the dimension of antecedent variables $\mathbf{x} = (x_1, x_2, \cdots x_d)^T$, and $\tilde{A}_{i1}, \cdots, \tilde{A}_{id}$ denotes fuzzy sets defined for the respective antecedent variable. The rule consequent is represented by $[y_{il}, y_{ir}]$ which corresponds to the centroid of the type-2 interval consequent set. Parameter c is the number of fuzzy rules. Assuming that y_{il} and y_{ir} have been rule-reordered [16], then the decision function in terms of interval type-2 fuzzy model by fuzzy-mean defuzzification is

$$[y_l, y_r] = \frac{\sum_{i=1}^{c} \mu_{\tilde{A}_i}(\mathbf{x})[y_{il}, y_{ir}]}{\sum_{i=1}^{c} \mu_{\tilde{A}_i}(\mathbf{x})}, \tag{2}$$

$$F(\mathbf{x}) = \frac{1}{2}\left(y_l + y_r\right). \tag{3}$$

where $\mu_{\tilde{A}_i}(\mathbf{x})$ denotes the membership grade of $\mathbf{x}$ in the interval type-2 fuzzy set $\tilde{A}_i$, and is given by

$$\mu_{\tilde{A}_i}(\mathbf{x}) = [\underline{f}_i(\mathbf{x}), \overline{f}_i(\mathbf{x})]. \tag{4}$$

Here,

$$\overline{f}_i(\mathbf{x}) = \overline{u}_i(x_1) \bullet \cdots \bullet \overline{u}_i(x_d), \tag{5}$$

$$\underline{f}_i(\mathbf{x}) = \underline{u}_i(x_1) \bullet \cdots \bullet \underline{u}_i(x_d). \tag{6}$$

where $\bullet$ denotes t-norm, $\overline{u}_i(x_j)$ and $\underline{u}_i(x_j)$ are upper and lower membership functions of $\tilde{A}_{ij}$, $j = 1, \cdots, d$.

(2) and (3) could be calculated by

$$y_l = \frac{\sum_{i=1}^{L} \overline{f}_i(\mathbf{x}) y_{il} + \sum_{i=L+1}^{c'} \underline{f}_i(\mathbf{x}) y_{il}}{\sum_{i=1}^{L} \overline{f}_i(\mathbf{x}) + \sum_{i=L+1}^{c'} \underline{f}_i(\mathbf{x})}, \tag{7}$$

$$y_r = \frac{\sum_{i=1}^{R} \underline{f}_i(\mathbf{x}) y_{ir} + \sum_{i=R+1}^{c'} \overline{f}_i(\mathbf{x}) y_{ir}}{\sum_{i=1}^{R} \underline{f}_i(\mathbf{x}) + \sum_{i=R+1}^{c'} \overline{f}_i(\mathbf{x})}. \quad (8)$$

Here, R and L are switch points which can be obtained by KM algorithm [8]. The identification of parameters is to minimize the error measure

$$e = \frac{1}{2}\left[F(\mathbf{x}) - y\right]^2 \quad (9)$$

where $F(\mathbf{x}) = (y_l + y_r)/2$.

As Mendel claimed in [8], the model (2) and (3) can effectively describes the relationship between input and output under the uncertainty.

2.2 RVM

SVM which possesses high generalization ability has been found to be robust in many applications [17]. However, one can easily show examples or experiments where optimal generalization performances of SVM are achieved with the number of support vectors more or less than 50% training samples. If we tune hyperparameters to reduce the number of support vectors, then the desirable generalization performances will loss [18]. In other words, when we extract support vectors of the ordinary solution of SVM for generating fuzzy rules, the optimal model selection procedures will produce too many fuzzy rules, which will result some redundancy rules in rule base, and will confuse our understandings. This phenomenon indicates that fuzzy modeling based on support vector learning will lead to a tradeoff problem between transparency in rule-base fuzzy system and performances [6].

RVM could be seen as a simlified version of SVM, since it tries to approximate SVM solutions by another comprised by a much smaller number of reduced-set vectors. At the same time, it keeps the acceptable performances of the original SVM solutions. For the problem of decision function approximation, RVM is formulated as [18]

$$F(\mathbf{x}) = \sum_{i=1}^{c'} \theta_i' k(\mathbf{x}, \mathbf{z}_i) + b. \quad (10)$$

where c' denotes the number of reduced-set vectors; the reduced set consists of $\{(\theta_i', \mathbf{z}_i)\}_{i=1,\cdots c'}$, and $\mathbf{z}_i$ denotes reduced-set vector. The usual way to construct reduce set is to minimize

$$\min_{\theta_i, \mathbf{z}_i} \left\| \sum_{i=1}^{c} \theta_i \Phi(\mathbf{x}_i) - \sum_{i=1}^{c'} \theta_i \Phi(\mathbf{z}_i) \right\|, \quad (11)$$

where θ_i and $\mathbf{x}_i$ are the original solutions of SVM; c denotes the number of support vectors; $k(\mathbf{x}, \mathbf{z}_i) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{z}_i))$. Several optimal methods had been proposed to minimize (10) [18], [19].

2.3 Combination

The aim of this paper is to integrate RVM to IT2FLS, and lead to a network which can use the advantages of each technique offers. In order to achieve the combination, it is useful to examine the similarities and differences between them.

Development of IT2FLS combined with RVM is to extract reduced-set vectors for generating interval type-2 fuzzy rules, so c and c' in (2) and (10) are equal. Therefore, one reduced-set vector corresponds to one rule. In this way, the model structure is automatically determined by reduced-set vectors. To accomplish it as the type-1 Takagi-Sugeno fuzzy model with RVM [6], the implementation of kernels also needs to be available for the interval type-2 fuzzy memberships. In next section, we will discuss a new concept, interval kernel.

3 Interval kernel

Chen *et al*. [2] firstly introduced the reference function $\mu(x)$ to construct the kernels.

$$k(\mathbf{x}, \mathbf{z}_i) = \prod_{j=1}^{d} \mu(x_j, z_{ij}, \Theta_j). \quad (12)$$

Clearly, the reference function can be regarded as exact symmetrical type-1 fuzzy membership. It corresponds to the ordinary fuzzy set A_{ij}, and Θ_j denotes the size of support of A_{ij}, z_{ij} denotes the center of it. Thus, (12) can be regarded as product type multidimensional membership function, and the denotation $\prod$ is also considered as a fuzzy logical operator which is *t*-norm-based algebra product. Hence, it is believed that the resulting product-type kernels are the form of type-1, and they are "exact".

However, due to the known disadvantages of type-1 logic, it seems doubtful whether these "exact" kernels are adequate to cope with uncertain or complicated real data. Hence, we generalize ordinary Mercer kernel onto interval. This leads to the concept of interval kernel,

Definition (interval kernel): If $\Theta \in \Omega$ are kernel parameters of ordinary Mercer kernel, an interval kernel is a function $k^I : X \times X \to IR$ for all $\mathbf{x}, \mathbf{z} \in X \subseteq R^d$, i.e.,

$$k^I(\mathbf{x}, \mathbf{z}) = [\underline{k}(\mathbf{x}, \mathbf{z}), \overline{k}(\mathbf{x}, \mathbf{z})], \quad (13)$$

where

$$\underline{k}(\mathbf{x}, \mathbf{z}) = \inf_{\Theta \in \Omega} k(\mathbf{x}, \mathbf{z}), \quad (14)$$

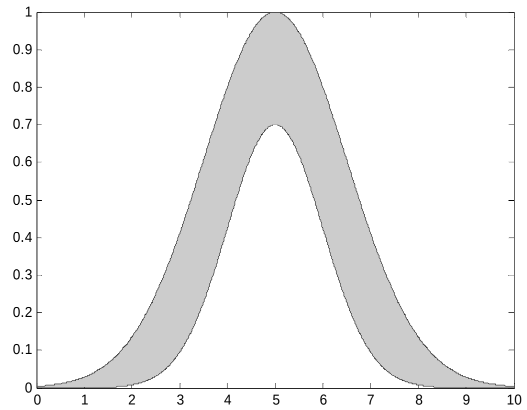


Fig. 1. Symmetric FOU – Gaussian UMF and scaled Gaussian LMF

$$\overline{k}(\mathbf{x},\mathbf{z}) = \sup_{\Theta\in\Omega} k(\mathbf{x},\mathbf{z}). \quad (15)$$

The expression (14) and (15) respectively represent the upper and lower kernel function of interval kernel. Obviously, both of them are still ordinary Mercer kernel. The following proposition can be viewed as showing that interval kernels satisfy a number of closure properties, allowing us to create more complicated interval kernels from simple building blocks.

Proposition 1 (Closure properties): Let k_1^I and k_2^I be interval kernels over $X \times X$, $a \in R^+$. Then the following functions are interval kernels:

(i) $k^I(\mathbf{x},\mathbf{z}) = k_1^I(\mathbf{x},\mathbf{z}) + k_2^I(\mathbf{x},\mathbf{z})$,

(ii) $k^I(\mathbf{x},\mathbf{z}) = ak_1^I(\mathbf{x},\mathbf{z})$,

(iii) $k^I(\mathbf{x},\mathbf{z}) = k_1^I(\mathbf{x},\mathbf{z})k_2^I(\mathbf{x},\mathbf{z})$.

The proofs are easy, and leave them to readers.

Now, assuming (14) and (15) are constructed by (12), then (13) could be particularly formularized as

$$\begin{aligned} k^I(\mathbf{x},\mathbf{z}_i) &= \prod_{j=1}^{d} [\underline{\mu}(x_j, z_{ij}, \underline{\Theta}_j), \overline{\mu}(x_j, z_{ij}, \overline{\Theta}_j)] \\ &= [\prod_{j=1}^{d} \underline{\mu}(x_j, z_{ij}, \underline{\Theta}_j), \\ &\quad \prod_{j=1}^{d} \overline{\mu}(x_j, z_{ij}, \overline{\Theta}_j)]. \end{aligned} \quad (16)$$

According to the theorem 3 in [9], (16) could reappear as (17) where $\bullet$ denotes product *t*-norm; $\underline{\mu}(x_j, z_{ij}, \underline{\Theta}_j)$ and $\overline{\mu}(x_j, z_{ij}, \overline{\Theta}_j)$ are upper membership function (UMF) and lower membership function (LMF) which bound the footprint of uncertainty (FOU) of an interval type-2 membership function $\mu_{\tilde{A}_{ij}}$. In this way, it is clear that the interval kernel is equal to primary membership grades of an interval type-2 membership function. Here, we provide a special FOU for constructing interval kernel, i.e., symmetric FOU – Gaussian UMF and scaled Gaussian LMF [19].

$$\overline{\mu}(x) = \exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\}, \quad (18)$$

$$\underline{\mu}(x) = s\exp\left\{-\frac{1}{2}\left(\frac{x}{\lambda\sigma}\right)^2\right\}. \quad (19)$$

Its a realization is shown in Fig.1.

Thus, by using the interval kernel (16) instead of ordinary Mercer kernel, the decision function in terms of an interval RVM model for crisp input and output is formulated as

$$[\underline{F}(\mathbf{x}), \overline{F}(\mathbf{x})] = \sum_{i=1}^{c'} \theta_i' k^I(\mathbf{x},\mathbf{z}_i) + [b,b]. \quad (20)$$

$$F(\mathbf{x}) = \left\{\underline{F}(\mathbf{x}) + \overline{F}(\mathbf{x})\right\}/2. \quad (21)$$

Note that (21) employs mean computation which looks like the form of defuzzified output of an interval type-2 fuzzy model. Moreover, this model differs from traditional interval regression model [21], since it directly exploits interval kernel other than interval coefficients to construct regression model. It is apparent that the model presented is more simple and interpretable. Indeed, if the mean computation was abandoned, (20) will directly describe a relationship between crisp input and interval output.

4 RV-based IT2FNN

Based on above disscussions, a new model named reduced-set vector-based interval type-2 fuzzy neural network, i.e., RV-based IT2FNN is formulated in this section. The goal of the RV-based IT2FNN is to construct a good neuro-fuzzy model that has a good structure, good knowledge interpretability, high prediction accuracy, and has good generalization capability with a small number of rules. Thus, a hybrid learning mechanism is presented to achieve it. It involves two sub-algorithms: bottom-up simplification algorithm and ε insensitive learning algorithm.

4.1 Model formulation

By interval arithmetic, we unify (20) and (21) as more visible expression

$$F(\mathbf{x}) = \sum_{i=1}^{c'} \theta_i' k^h(\mathbf{x},\mathbf{z}_i) + b. \quad (22)$$

where the hybrid kernel is defined as

$$k^h(\mathbf{x},\mathbf{z}_i) = \underline{k}(\mathbf{x},\mathbf{z}_i)/2 + \overline{k}(\mathbf{x},\mathbf{z}_i)/2. \quad (23)$$

$$k^I(\mathbf{x},\mathbf{z}_i) = [\sup_{x\in X}\int_{X_1}\cdots\int_{X_d} \underline{\mu}(x_1, z_{i1}, \underline{\Theta}_1)\bullet\cdots\bullet\underline{\mu}(x_d, z_{id}, \underline{\Theta}_d)/x,\ \sup_{x\in X}\int_{X_1}\cdots\int_{X_d} \overline{\mu}(x_1, z_{i1}, \overline{\Theta}_1)\bullet\cdots\bullet\overline{\mu}(x_d, z_{id}, \overline{\Theta}_d)/x]. \quad (17)$$

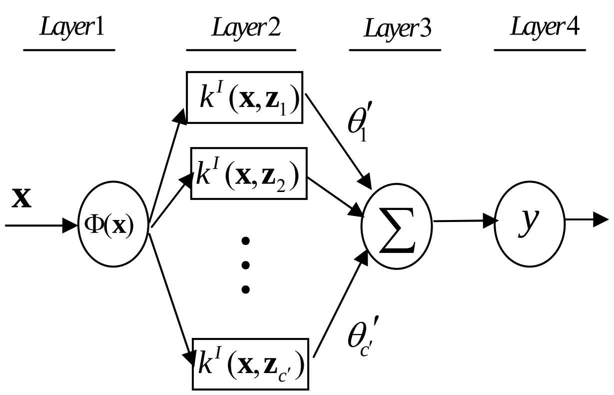


Fig. 2. The architecture of RV-based IT2FNN

Thus, the interval RVM model (20) and (21) are integrated into a standard RVM model

Actually, hybrid kernel which is encoded as non-negative combination coefficients and admissible Mercer kernels had ever been proposed to make the resulting kernels more flexible so as to accommodate various requirements [22], [23]. This seems very similar to the reason that interval type-2 fuzzy memberships are presented. In this paper, we employ RVM model (22) to produce reduced-set vectors, and then extract them to generate interval type-2 fuzzy rules for constructing RV-based IT2FNN.

We utilize the definition of interval kernel (16), (18) and (19) to construct hybrid kernel employed in RVM model. As a result, the learning of RVM model leads to a reduced set $\{(\theta_i', \mathbf{z}_i)\}_{i=1,\cdots c'}$ and the basic structure of hybrid kernel, i.e.,

$$\underline{k}(\mathbf{x},\mathbf{z}_i) = s^d \exp\left\{\sum_{j=1}^{d}\left(-\frac{1}{2}\left(\frac{x_j - z_{ij}}{\lambda\sigma_{ij}}\right)^2\right)\right\} \quad (24)$$

$$\overline{k}(\mathbf{x},\mathbf{z}_i) = \exp\left\{\sum_{j=1}^{d}\left(-\frac{1}{2}\left(\frac{x_j - z_{ij}}{\sigma_{ij}}\right)^2\right)\right\} \quad (25)$$

Based on these information, reduced-set vectors are choosen as the centers of interval type-2 fuzzy memberships, and then interval type-2 if-then fuzzy rules are directly generated. The membership grade of $\mathbf{x}$ in the interval type-2 fuzzy set $\tilde{A}_i$ is

$$\mu_{\tilde{A}_i}(\mathbf{x}) = [s^d \exp\left\{\sum_{j=1}^{d}\left(-\frac{1}{2}\left(\frac{x_j - z_{ij}}{\lambda\sigma_{ij}}\right)^2\right)\right\}, \exp\left\{\sum_{j=1}^{d}\left(-\frac{1}{2}\left(\frac{x_j - z_{ij}}{\sigma_{ij}}\right)^2\right)\right\}]. \quad (26)$$

Thus, the primary membership grades of antecedent interval type-2 fuzzy set $\tilde{A}_{ij}$ is

$$\underline{\mu}_{\tilde{A}_{ij}}(\mathbf{x}) = s\exp\left\{-\frac{1}{2}\left(\frac{x_j - z_{ij}}{\lambda\sigma_{ij}}\right)^2\right\};$$

$$\overline{\mu}_{\tilde{A}_{ij}}(\mathbf{x}) = \exp\left\{-\frac{1}{2}\left(\frac{x_j - z_{ij}}{\sigma_{ij}}\right)^2\right\}. \quad (27)$$

Hence, a new model named RV-based IT2FNN is created, and its network is shown in Fig.2.

$$[\underline{F}(\mathbf{x}), \overline{F}(\mathbf{x})] = \sum_{i=1}^{c'} \theta_i' \mu_{\tilde{A}_i}(\mathbf{x}) + [b, b]. \quad (28)$$

$$F(\mathbf{x}) = \left\{\underline{F}(\mathbf{x}) + \overline{F}(\mathbf{x})\right\}/2. \quad (29)$$

The resulting model consists of a reduced set of linguistic rules in the following form:

$$R_i : \textit{If } x_1 \textit{ is } \tilde{A}_{i1} \textit{ and } x_2 \textit{ is } \tilde{A}_{i2} \textit{ and} \cdots x_d \textit{ is } \tilde{A}_{id},$$
$$\textit{then} \quad y_i = F(\mathbf{z}_i) \quad \textit{for } i = 1, 2 \cdots, c \quad (30)$$

where $\mathbf{z}_i$ denotes the reduced-set vector obtained.

Noticeably, the rules are different from the traditional interval type-2 fuzzy rules, here the consequent is represented as a singlton not the centroid of the type-2 interval consequent set However, it does not violate the nature of interval type-2 fuzzy logic system, since a fuzzy logic system is type-2 as long as any one of its antecedent or consequent sets is type-2 [8].

For the model proposed, we directly choose interval kernel as interval type-2 fuzzy memberhip function, and the model can be represented as series expansion of interval kernel. It means that traditional type-reduction procedure in interval type-2 fuzzy inference system is given up, and defuzzification procedure is successively implemented after inference. As Chiang [4] claimed that the remove of denominator of the fuzzy basis function does not also violate the spirit of fuzzy inference system.

Moreover, the proposed model could also be regarded as a four-layer network. For layer1, it maps the input from input space to a higher dimensional feature space. This layer is only for clarity and is not needed in actual computation. Each node in layer 2 represents an interval kernel which is seen as interval type-2 fuzzy membership, and each reduced-set vector corresponds one interval type-2 fuzzy rule. The single node in layer3 is a fixed node, which performs the function of overall aggregation of all the interval type-2 fuzzy rules with weighted paramter θ_i'. Layer4 is a defuzzification unite. It is easy to find out that the proposed RV-based IT2FNN has adaptive structure according to reduced-set vectors, while the structure of traditional interval T2FNN needs to be specified in advanced. In next subsection, hybrid learning mechanism is proposed to training the RV-based IT2FNN.

4.2 Hybrid learning mechanism

Given n training data pairs $\{(\mathbf{x}_k, y_k)\}_{k=1,\cdots,n} \subset R^d \times R$ with unknown joint distribution $P(\mathbf{x}, y)$, a hybrid learning mechanism which involves two sub-algorithms is presented for identifying the structure and parameters of RV-based IT2FNN.

Firstly, the bottom-up simplification algorithm is utilized to learn (22) with hybrid kernel, and then rapidly produce reduced-set vectors for construct the structure of RV-based IT2FNN. Thus, the rule number of RV-based IT2FNN is entirely determined by the number of reduced-set vectors, and it need not to fix on the rule number in advance.

Here, the bottom-up simplification algorithm can be viewed as a bottom-up hierarchical clustering procedure, in which two nearest support vectors are iteratively selected belonging to the same class and replaced by a newly constructed one. Moreover, the construction of the new vectors requires to find the unique maximum point of a one-variable function on (0,1), not to minimize a function of many variables with local minima [19]. In our paper, an additional proposition is given in the followings for the hybrid kernel created by (24) and (25),

Proposition 2: For the hybrid kernel

$$k^h(\mathbf{x},\mathbf{z}) = \frac{1}{2}\exp\left\{\sum_{j=1}^{d}\left(-\frac{1}{2}\left(\frac{x_j - z_j}{\sigma_j}\right)^2\right)\right\} + \frac{1}{2}s^d \exp\left\{\sum_{j=1}^{d}\left(-\frac{1}{2}\left(\frac{x_j - z_j}{\lambda\sigma_j}\right)^2\right)\right\},$$

the 2-norm optimal approximation of $M = m\Phi(\mathbf{x}_i) + (1-m)\Phi(\mathbf{x}_j)$, $m = \alpha_i / (\alpha_i + \alpha_j)$, $\alpha_i\alpha_j > 0$, is the image of input vector $\mathbf{y}$ determined by

$$\mathbf{z} = k\mathbf{x}_i + (1-k)\mathbf{x}_j \tag{31}$$

where k is the maximum point of

$$g(k) = mC_{ij}^{(1-k)^2} + (1-m)C_{ij}^{k^2} \tag{32}$$

with $C_{ij} = k^h(\mathbf{x}_i, \mathbf{x}_j)$.

According to the addtional proposition for hybrid kernel, the overall simplification bottom-up algorithm used here is as the same as the one in [19]. So, after the learning of the first sub-algorithm, the structure of RV-based IT2FNN is directly created by following the procedures in section 3.1.

Subsequently, the second sub-algorithm is used to refine RV-based IT2FNN, i.e., ε insensitive learning algorithm. This algorithm has been developed for type-1 fuzzy modeling [6], [24]. Here, it is employed to tune the weighted parameters of our RV-based IT2FNN.

More specially, combine (28) and (29) that lead to

$$F(\mathbf{x}) = \sum_{i=1}^{c'} \theta_i' Med[\mu_{\tilde{A}_i}(\mathbf{x})] + b. \tag{33}$$

Let $\psi(\mathbf{x}_k) = \left[Med[\mu_{\tilde{A}_1}(\mathbf{x}_k)], \cdots, Med[\mu_{\tilde{A}_{c'}}(\mathbf{x}_k)]\right]^T$ and $\theta' = [\theta_1', \cdots, \theta_{c'}']^T$, then rewrite (33) for kth training data

$$F(\mathbf{x}_k) = \psi(\mathbf{x}_k)^T \theta' + b. \tag{34}$$

Using ε insensitive learning loss function, the learning algorithm has the following form [24]

$$\min_{\theta' \in R^{c'}, b \in R} E(\theta', b) = C\sum_{k=1}^{n}\left|y_k - \psi(\mathbf{x}_k)^T\theta' - b\right|_{\varepsilon} + \frac{1}{2}\theta'^T\theta' \tag{35}$$

The above minimization can be found by means of methodogy introduced by Vapnik [25]. Generally, the key idea is to construct a Lagrange function from the objective function by introducing a dual set of variables.

By introducing slack variables $\xi_k^+, \xi_k^- \geq 0$ for all data pairs we can write

$$\begin{aligned} &y_k - \psi(\mathbf{x}_k)^T\theta' - b \leq \varepsilon + \xi_k^+, \\ &-y_k + \psi(\mathbf{x}_k)^T\theta' + b \leq \varepsilon + \xi_k^-. \end{aligned} \tag{36}$$

Using (36), criterion (35) takes the form

$$E(\theta', b) = C\sum_{k=1}^{n}\left(\xi_k^+ + \xi_k^-\right) + \frac{1}{2}\theta'^T\theta'. \tag{37}$$

The minimization of (37) subject to constraints (36) and $\xi_k^+, \xi_k^- \geq 0$ leads to the following Lagrangian function (38) where $\lambda_k^+, \lambda_k^-, \mu_k^+, \mu_k^- \geq 0$ are Lagrange multipliers. The objective is to minimize the above Lagrangian with respect to θ', b, ξ_k^+, ξ_k^- and maximize with respect to the Lagrange multipliers. The following conditions for optimality are obtained by differentiating (20) with respect to θ', b, ξ_k^+,

$$G = \frac{1}{2}\theta'^T\theta' + C\sum_{k=1}^{n}\left(\xi_k^+ + \xi_k^-\right) - \sum_{k=1}^{n}\lambda_k^+\left(\varepsilon + \xi_k^+ - y_k + \psi(\mathbf{x}_k)^T\theta' + b\right) - \sum_{k=1}^{n}\lambda_k^-\left(\varepsilon + \xi_k^- + y_k - \psi(\mathbf{x}_k)^T\theta' - b\right) - \sum_{k=1}^{n}\left(\mu_k^+\xi_k^+ + \mu_k^-\xi_k^-\right) \tag{38}$$

ξ_k^- and setting the results equal to zero

$$\frac{\partial G}{\partial \theta'} = \theta' - \sum_{k=1}^{n}\left(\lambda_k^+ - \lambda_k^-\right)\psi(\mathbf{x}_k) = 0,$$
$$\frac{\partial G}{\partial b} = \sum_{k=1}^{n}\left(\lambda_k^+ - \lambda_k^-\right) = 0,$$
$$\frac{\partial G}{\partial \xi_k^+} = C - \lambda_k^+ - \mu_k^+ = 0,$$
$$\frac{\partial G}{\partial \xi_k^-} = C - \lambda_k^- - \mu_k^- = 0. \tag{39}$$

The last two conditions (21) and the requirements μ_k^+, $\mu_k^- \geq 0$ imply that $\lambda_k^+, \lambda_k^- \in [0, C]$. Thus, the weighted paramters of RV-based IT2FNN are estimated by

$$\theta' = \sum_{k=1}^{n}\left(\lambda_k^+ - \lambda_k^+\right)\psi(\mathbf{x}_k) \tag{40}$$

By inserting conditions (39) into Lagrangian (38) we obtain

$$G = -\frac{1}{2}\sum_{k=1}^{n}\sum_{j=1}^{n}\left(\lambda_k^+ - \lambda_k^-\right)\left(\lambda_j^+ - \lambda_j^-\right)\psi(\mathbf{x}_k)^T\psi(\mathbf{x}_k)$$
$$-\varepsilon\sum_{k=1}^{n}\left(\lambda_k^+ + \lambda_k^-\right) + \sum_{k=1}^{n} y_k\left(\lambda_k^+ - \lambda_k^-\right). \tag{41}$$

The maximization of (41) with respect to λ_k^+, λ_k^- subjec to constraints:

$$\sum_{k=1}^{n}\left(\lambda_k^+ - \lambda_k^-\right) = 0,$$
$$\lambda_k^+, \lambda_k^- \in [0, C] \tag{42}$$

is the Wolfe dual formulation of (38).

Hence, based on the above discussion, the overall hybrid learning mechanism for RV-based IT2FNN is summerized in the followings.

Hybrid learning mechanism of RV-based IT2FNN:

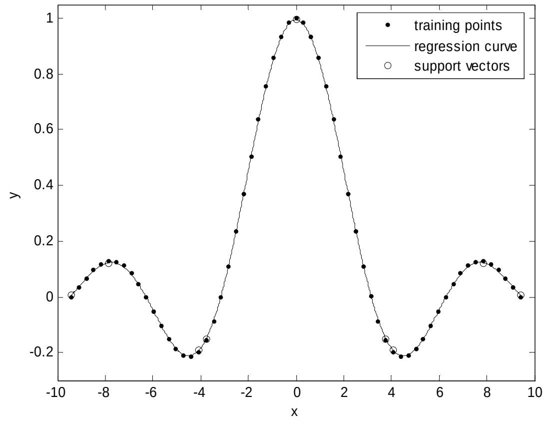
Initialization: interval kernel parameters; insensitive error ε_1 and regularization parameter C_1 for RVM ; insensitive error ε_2 and regularization parameter C_2 for insensitive learning; threshold of maximum marginal difference (MMD); training data; learning rate α .

Step 1) use (22) to produce reduced-set vectors according to bottom-up simplification algorithm;

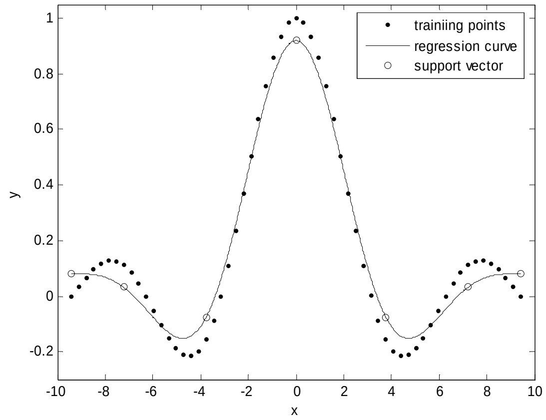
Step 2) construct RV-based IT2FNN on the basis of (24)- (29);

Step 3) adjust the weighted parameter θ' by ε insensitive learning algorithm;

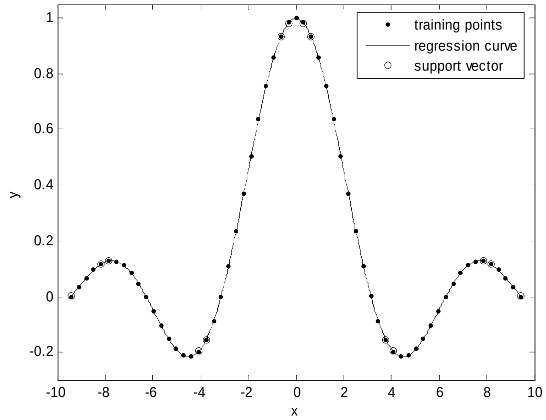
5 Simulations



(a) 9 support vectors with ε =0.005



(b) 7 support vectors with ε =0.08



(c) 14 support vectors with ε =0.001

Fig.3 Fuzzy modeling results using support vector

To illustrate the approach, we will design two simulations in the followings.

5.1 Sinc function

Let us consider the 1-D sinc function,

$$\text{sinc}(x) = \begin{cases} \sin(x)/x, & if\ x \neq 0 \\ 1 \quad , & if\ x = 0 \end{cases}. \tag{50}$$

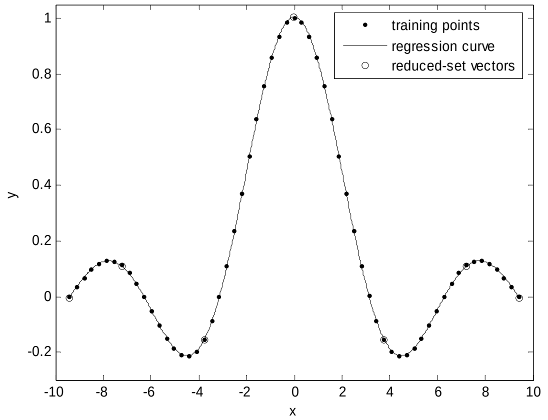


Fig.4 RV-based IT2FNN modeling results with 7

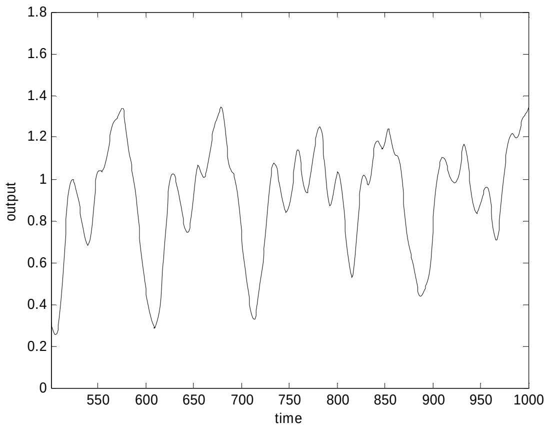
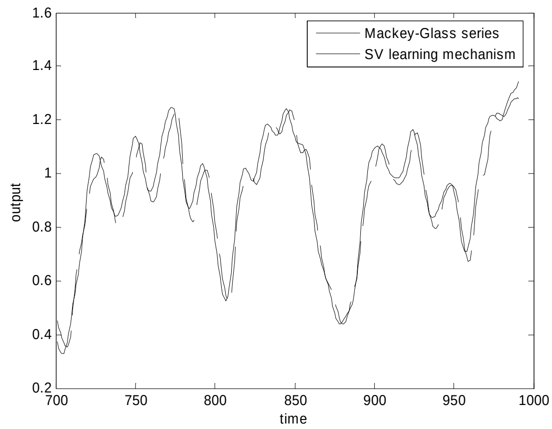


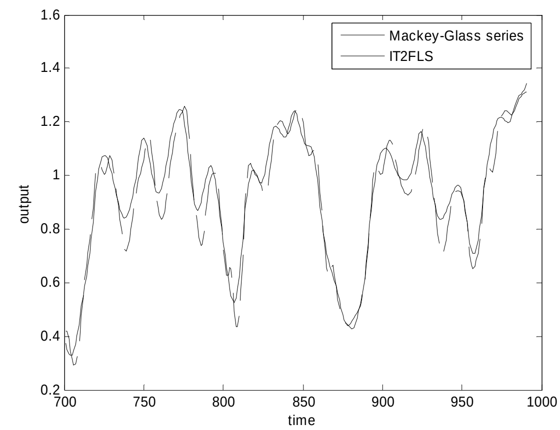
Fig. 5 Mackey-Glass chaotic time series

The training set consists of 61 training points which are generated range from -3π to 3π. Type-1 fuzzy model based on support vector learning mechanism learning mechanism with different ε for has been developed to fit this target [4]. In result, 9 support vectors are obtained with $C_1 = 10$, $\varepsilon_1 = 0.005$, $\sigma = 2.22$. Furthermore, we will discuss the role of parameter ε_1 which usually decides the number of support vectors used to construct the number of fuzzy rule. Here, parameter ε_1 is respectively reset to 0.08 and 0.001 with 7 support vectors obtained and 14 ones. All the results are shown in Fig.3. Obviously, high accuracy needs more number of rules, whereas fewer number of rules make the resulting model loss the nice performance.

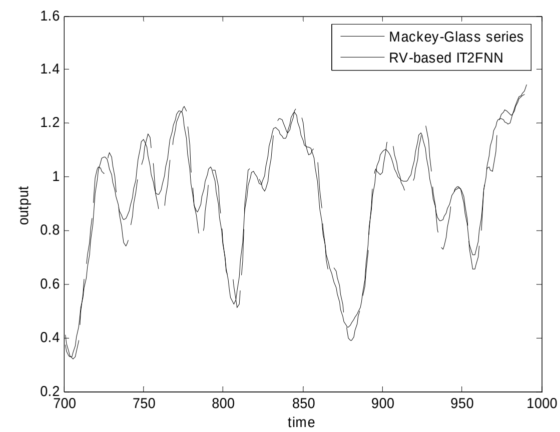
Our goal is to train the RV-based IT2FNN for fitting the desired target (50). For initial parameters, we set $C_1 = 10$, $\varepsilon_1 = 0.08$, $\sigma = 2.22$, MMD = 0.05, $\alpha = 0.2$, $s = 0.6$, $\lambda = 1.5$, $C_2 = 10$, $\varepsilon_2 = 0.01$. After hybrid learning, the regression curve of RV-based IT2FNN is shown in Fig.4. The training rms error is 0.0031 which is even smaller than 0.0036 obtained



(a) type-1 fuzzy model based on SV learning mechanism with 21 rules



(b) IT2FLS combined back propagation with 10 rules



(c) RV-based IT2FNN with 10 rules

Fig. 6 Prediction of Mackey-Glass chaotic time series with different fuzzy modeling

from type-1 fuzzy model with 9 support vectors. Consequently, the RV-based IT2FNN proposed could effectively decrease the number of fuzzy rules, at the same time possess acceptable performance.

Table.1 RASE for training and testing with different fuzzy modeling

	type-1 fuzzy model based on SV learning mechanism	IT2FLS combined back propagation method	RV-based IT2FNN with hybrid learning mechanism
RASE for training	0.0469	0.0294	0.0334
RASE for testing	0.0557	0.0610	0.0556
Number of rules	21	10	10

5.2 Prediction of Mackey-Glass time series

The Mackey-Glass chaotic time series can be represented as [4],[8]

$$\frac{ds(t)}{dt}=\frac{0.2s(t-\tau)}{1+s^{10}(t-\tau)}-0.1s(t). \quad (51)$$

When $\tau>17$, (51) exhibits chaotic behavior. In simulating (51), we converted it to a discrete-time equation by using Euler's method. Denoting

$$f(s,n)=\frac{0.2s(n-\tau)}{1+s^{10}(n-\tau)}-0.1s(n), \quad (52)$$

then

$$s(n+1)=s(n)+f(s,n), \quad (53)$$

where the initial value of $s(n)$ is set randomly. In our simulation, τ is chosen as 30.

The goal of this task is to use known values of the time series up to the point $s(n)$ to predict the value at some point in the future $s(n+P)$ .The standard method for this type of prediction is to create a mapping from D points of the time series spaced Δ apart, that is, $\left(s\left(n-(D-1)\Delta\right),\cdots,s\left(n-\Delta\right),s\left(n\right)\right)$, to a predicted future value $s(n+P)$. Here, the values $D=9$ and $\Delta=P=1$ are used, i.e., nine points values in the series are used to predict the value of the next time point. Fig. 5 shows 500 points, $s(501)-s(1000)$, of this chaotic series used to train and test our model.

Particularly, the 200 points of the series from $s(501)-s(700)$ are used as training data, i.e., 191 training patterns. Similarly, the final 300 points of the series from $s(701)-s(1000)$ are used as test data, i.e., 291 testing patterns.

In this simulation, we employ type-1 fuzzy model based on support vector learning mechanism [4] and IT2FLS combined back propagation method [9] to predict Mackey-Glass time series, and compare them with proposed RV-based IT2FNN.

Here, we use the standard deviation of time series to estimate the width parameter σ of all fuzzy models. The other parameters of type-1 fuzzy model based on support vector learning mechanism are set to $C_1=10$, $\varepsilon_1=0.07$. For IT2FLS combined back propagation, we initialize the other parameters randomly with 10 fuzzy rules. In our model, the initialization is set to $C_1=10$, $\varepsilon_1=0.07$, $\alpha=0.2$, $\text{MMD}=0.01$, $s=0.9$, $C_2=0.5$, $\varepsilon_2=0.02$, $\lambda=1.35$. After respective learning, the results of prediction with different fuzzy modeling are shown in Fig. 6. Moreover, the measure, root average squared error (RASE), in the followings is used to evaluate the predictive accuracy

$$\text{RASE}=\sqrt{\frac{\sum_{i=1}^{n}\left(y_k-F(\mathbf{x}_k)\right)^2}{n}}. \quad (54)$$

The results of RASE for training and testing are shown in Table.1.

Clearly, similar with the results of section 5.1, the approach of type-1 fuzzy model based on SV learning mechanism leads to more fuzzy rules. Besides, the back propagation method makes the IT2FLS overfit the training patterns, and the reason might be the bad guess about the initial parameter locations or initial structure of IT2FLS. The predictive accuracy of RV-based IT2FNN is acceptable. It is of great important that the number of fuzzy rules in proposed model is small. This makes the rule base more transparent and interpretable.

6 Conclusion

This paper has tried to build a structural framework for interval type-2 fuzzy modeling, i.e., RV-based IT2FIS with hybrid learning machanism. In the first instance, the functional equivalence between interval kernel and interval type-2 membership function is established. In this way, the interval type-2 membership function can be ceated by means of kernel transformation in statistical learning theory. Subsequently, interval type-2 fuzzy rules could be rapidly generated according to the reduced-set vectors which are obtained from RVM with interval kernel by using the bottom-up simplification algorithm. The resulting model in forms of these interval type-2 fuzzy rules gives up traditional type reduction procedure, and it is regarded as a four layers neural network. It is of great importance that RV-based IT2FIS possess adaptive structure which is directly determined by reduced-set vectors, and it need not to initialize the structure of IT2FLS in advance. Thus, RV-based IT2FNN with hybrid learning mechanism effectively complishes the structure identification of interval type-2 fuzzy sysstem and makes it more accurate. In result, the fuzzy model proposed preserves advantages of both RVM and IT2FLS.

References:

[1] S. Mitra, Y. Hayashi, Neuro-Fuzzy Rule Generation: Survey in Soft Computing Framework, *IEEE Transactions on Neural Network*, Vol.11, No.3, 2000, pp. 748-768.

[2] Y. Chen, J. Z. Wang, Support Vector Learning for Fuzzy Rule-based Classification Systems, *IEEE Transactions on Fuzzy Systems*, Vol.11, No.6, 2003, pp. 716-728.

[3] S. Papadimitriou, K. Terzidis, S. Mavroudi, Fuzzy Rule Based Classifiers from Support Vector Learning. *WSEAS Transactions on Computers*, Vol.4, No.7, 2005, pp. 661-670.

[4] J.H. Chiang, P.Y. Hao, Support Vector Learning Mechanism for Fuzzy Rule-based Modeling: A New Approach, *IEEE Transactions on Fuzzy Systems*, Vol.12, No.1, 2004, pp. 1-12.

[5] J. Shen, Yu-Ru Syau, E.S. Lee, Support Vector Fuzzy Adaptive Network in Regression Analysis. *Computers and Mathematics with Applications*, Vol.54, 2007, pp. 1353-1366.

[6] L. Yu, J. Xiao, L. Z. Liu, A New Reduced-set vector-based Takagi-Sugeno fuzzy model. *Kong Zhi Li Lun Yu Ying Yong*, 2009. (To be published)

[7] L. Yu, J. Xiao, Y. F. Bai, A New Fuzzy Modeling Approach Based on Support Vector Regression, *Fourth International Conference on Fuzzy Systems and Knowledge Discovery*, Vol.3, 2007, pp. 578-584.

[8] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Direction*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[9] Q. L. Liang, J. M. Mendel, Interval Type 2 Fuzzy Logic Systems: Theory and Design, *IEEE Transactions on Fuzzy systems*, Vol.8, No.5, 2000, pp. 535-550.

[10] M. Singh, S. Srivastava, M. Hanmandlu, Identification and Control of Nonlinear Systems using Type-2 Fuzzy Set Based Neuro-fuzzy Model, *WSEAS Transactions on Computers*, Vol.6, No.6, 2007, pp. 935-940.

[11] M. Singh, S. Srivastava, J. R. P. Gupta, M. Hanmandlu, Extraction of System Dynamics for a Nonlinear System using Type-2 Fuzzy Sets Based Neuro-fuzzy Model, *WSEAS Transactions on Computers*, Vol.6, No.6, 2007, pp. 895-870.

[12] C. H. Wang, C. S. Cheng, T.T. Lee, Dynamical Optimal Training for Interval Type-2 Fuzzy Neural Network, *IEEE Transactions on Systems, Man, Cybern. B, Cybern*, Vol.34, No.3, 2004, pp. 1462-1477.

[13] G. M. Mendez, L. J. Ismael, Orthogonal-back Propagation Hybrid Learning Algorithm for Interval Type-1 Non-single Type-2 Fuzzy Logic Systems, *WSEAS Transactions on Systems*, Vol.4, No.3, 2005, pp. 212-218.

[14] G. M. Mendez, L. J. Ismael, First-order Interval Type-2 TSK Fuzzy Logic Systems using a Hybrid Learning Algorithm, *WSEAS Transactions on Computers*, Vol.4, No.4, 2005, pp. 378-384.

[15] Ö. Uncu, I. B. Türksen, Discrete Interval Type-2 Fuzzy System Models Using Uncertainty in Learning Parameters, *IEEE Transactions on Fuzzy Systems*, Vol.15, No.1, 2007, pp. 90-106.

[16] J. M. Mendel, Computing Derivatives in Interval Type-2 Fuzzy Logic Systems, *IEEE Transactions on Fuzzy Systems*, Vol.12, No.1, 2004, pp. 84-98.

[17] A. J. Smola, B. Schölkopf, *A Tutorial on Support Vector Regression*, NeuroCOLT2 Technical report, Royal Holloway College, 1998

[18] B. Schölkopf, A. J. Smola, *Learning with Kernels*. Cambridge, MA: MIT Press, 2002.

[19] D. D. Nguyen, T. B. Ho, A Bottom-up Method for Simplifying Support Vector Solutions, *IEEE Transactions on Neural Network*, Vol.17, No.3, 2006, pp.792-796.

[20] J. M. Mendel, H. W. Wu, Type-2 Fuzzistics for Symmetric Interval Type-2 Fuzzy Sets: Part1, Forward Problems, *IEEE Transactions on Fuzzy Systems*, Vol.14, No.6, 2004 , pp.781-792.

[21] C. Hwang, D. H. Hong, K. H. Seok, Support Vector Interval Regression Machine for Crisp Input and Output data, *Fuzzy Sets and Systems*, Vol.157, 2006, pp. 1114-1125.

[22] G. F. Smits, E. M. Jordaan, Improved SVM Regression using Mixtures of Kernels, *Proceedings of the International Joint Conference of Neural Network*, Vol.3, 2002, pp. 2785-2790.

[23] Y. Tan, J. Wang, A Support Vector Machine with a Hybrid Kernel and Minimal Vapnik-Chervonenkis Dimension. *IEEE Transactions on Knowledge and Data Engineering*, Vol.16, No.4, 2004, pp. 385-395.

[24] J. M. Łęski, ε Insensitive Fuzzy c-Regression Models: Introduction to ε Insensitive Fuzzy Modeling, *IEEE Transactions on Systems, Man, Cybern. B, Cybern*, Vol.34, No.1, 2004, pp. 4-15.

[25] Vapnik V. *Statistical Learning Theory*. Wiley New York, 1998.