

Possibilistic Regression Analysis by Support Vector Machine

Pei-Yi Hao

Department of Information Management.
National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan, R.O.C.
e-mail: haupy@cc.kuas.edu.tw

Abstract—Support vector machines (SVMs) have been very successful in pattern recognition and function estimation problems for crisp data. This paper proposes a new method to evaluate interval linear and nonlinear regression models combining the possibility and necessity estimation formulation with the principle of SVM. For data sets with crisp inputs and interval outputs, the possibility and necessity models have been recently utilized, which are based on quadratic programming approach giving more diverse spread coefficients than a linear programming one. The SVM also uses quadratic programming approach whose advantage in interval regression analysis is to be able to perform interval nonlinear regression analysis by constructing an interval linear regression function in a high dimensional feature space. The proposed algorithm is a attractive approach to modeling nonlinear interval data, and is model-free method in the sense that we do not have to assume the underlying model function for interval nonlinear regression model with crisp inputs and interval output. Experimental results are then presented which indicate the performance of this algorithm.

Keyword: *Interval regression analysis, possibility and necessity models, Support vector machines (SVMs), Support vector regression machines, Quadratic programming.*

I. INTRODUCTION

Regression analysis is to find a crisp relationship between the dependent and independent variables and to estimate a variance of measurement error. If a phenomenon under consideration is not governed by random variables but is governed by possibility variables, it is more natural to seek a fuzzy functional relationship for the given data being fuzzy or crisp. That is to say, the vague phenomenon where it is not assumed to follow randomness should be modeled by a fuzzy functional relationship. Fuzzy regression analysis can be classified into two categories; the first is possibilistic regression analysis based on possibility concepts and the second is least squares method for minimizing errors between the given outputs and the estimated outputs. Possibilistic regression derives new estimators by dealing directly with models formulated in a possibility context. Possibilistic regression analysis was first proposed by Tanaka et al. [1] where a fuzzy linear system was used as a regression model. Since membership functions of fuzzy sets are often described as possibility distributions, this approach is usually called possibilistic regression analysis [2,3].

Support vector machines (SVMs) have been very successful in pattern recognition and function estimation problems for crisp data [5,6]. Jeng *et al.* [7] first proposed the support vector interval regression networks (SVIRNs) for interval regression analysis by utilizing the ε -insensitive support vector regression (ε -SVR) in [5,6]. Hwang *et al.* [8] proposed the support vector interval regression machine (SVIRM) to evaluate the interval regression model combining the possibility estimation formulation and the property of central tendency with the principle of ε -SVR. Hao proposed a ν -support vector interval regression network (ν -SVIRN) to evaluate interval linear and nonlinear regression models for crisp input and output data [9]. As it is difficult to select an appropriate value of the insensitive tube width in ε -SVR, the ν -SVIRN alleviates this problem by utilizing a new parametric-insensitive loss function. The ν -SVIRN automatically adjusts a flexible parametric-insensitive zone of arbitrary shape and minimal size to include the given data. For a priori chosen ν , ν is an upper bound on the fraction of training errors and a lower bound on the fraction of support vectors. Hence, the selection of ν is more intuitive.

In this paper, we propose a new method to evaluate interval linear and nonlinear regression models combining the possibility and necessity estimation formulation with the principle of support vector learning approach. The SVM approach allows us to perform interval nonlinear regression analysis by constructing an interval linear regression function in a high-dimensional feature space. Hence, the proposed algorithm is a very attractive approach to model nonlinear interval data. This algorithm is model-free method in the sense that we do not have to assume the underlying model function for interval nonlinear regression model with crisp inputs and interval output. This model-free method turns out to be a promising method which has been attempted to treat interval nonlinear regression model.

II. A REVIEW OF PREVIOUS WORKS

A. Interval Regression Formulation Unifying the Possibility and Necessity Models

For a data set with crisp inputs and interval outputs, we can consider two interval regression models, i.e., the possibility and necessity models. In this section, we review the unified QP approach [4] to obtain the possibility and necessity models

simultaneously. In this unified approach, we assume for simplicity that the center coefficients of the possibility regression model and necessity regression model are same.

Suppose that we are given training data $\{(\mathbf{x}_1, Y_1), \dots, (\mathbf{x}_n, Y_n)\}$, where $\mathbf{x}_i = (x_{i1}, \dots, x_{im})^T$ is the i -th input vector, $Y_i = (y_i, e_i)$ is the corresponding interval output that consists of a center y_i and a radius e_i . For this data set, the possibility and necessity estimation models are denoted as

$$Y^*(\mathbf{x}_i) = A_0^* + A_1^* x_{i1} + \dots + A_m^* x_{im} \quad i = 1, \dots, n \quad (1)$$

$$Y_*(\mathbf{x}_i) = A_{*0} + A_{*1} x_{i1} + \dots + A_{*m} x_{im} \quad i = 1, \dots, n \quad (2)$$

where the interval coefficients A_i^* and A_{*i} are denoted as $A_i^* = (a_i^*, c_i^*)$ and $A_{*i} = (a_{*i}, c_{*i})$, respectively. The estimated interval $Y^*(\mathbf{x}_i)$ by the possibility model always includes the observed interval Y_i , whereas the estimated interval $Y_*(\mathbf{x}_i)$ by the necessity model should be included in the observed interval Y_i . We can denote the interval coefficients and as

$$A_i^* = (a_i, c_i + d_i) \text{ and } A_{*i} = (a_i, c_i)$$

which satisfies the condition $A_{*i} \subseteq A_i^*$, $i = 0, 1, \dots, m$, since c_i and d_i are assumed to be nonnegative. Therefore, by interval arithmetic the possibility model $Y^*(\mathbf{x}_i)$ and the necessity model $Y_*(\mathbf{x}_i)$ can be written as

$$Y^*(\mathbf{x}_i) = (\mathbf{a}'\mathbf{x}_i, \mathbf{c}'|\mathbf{x}_i| + \mathbf{d}'|\mathbf{x}_i|) \quad (3)$$

$$Y_*(\mathbf{x}_i) = (\mathbf{a}'\mathbf{x}_i, \mathbf{c}'|\mathbf{x}_i|) \quad (4)$$

The two objective functions in the possibility and the necessity models are assumed as follows:

$$\min J_P = \sum_{i=1}^n (\mathbf{c}'|\mathbf{x}_i| + \mathbf{d}'|\mathbf{x}_i|) : \text{Possibility model} \quad (5)$$

$$\max J_N = \sum_{i=1}^n \mathbf{c}'|\mathbf{x}_i| : \text{Necessity model} \quad (6)$$

The objective function of the unified model can be obtained by the combination of these two objective functions. The combination of two functions yields the following:

$$\min J_P - J_N = \sum_{i=1}^n \mathbf{d}'|\mathbf{x}_i|. \quad (7)$$

Then, the objective function in the unified approach by QP can be assumed as the following quadratic function:

$$J = \sum_{i=1}^n (\mathbf{d}'|\mathbf{x}_i|)^2 = \mathbf{d}' \left(\sum_{i=1}^n |\mathbf{x}_i|' |\mathbf{x}_i| \right) \mathbf{d} \quad (8)$$

Therefore, interval regression analysis [4] is to determine the interval coefficients A_i^* and A_{*i} , $i = 1, \dots, m$ that minimize the aforementioned objective function and satisfy inclusion

relations $Y^*(\mathbf{x}_i) \subseteq Y_i \subseteq Y_*(\mathbf{x}_i)$. This problem can be described as the following QP problem:

$$\begin{aligned} \min_{\mathbf{a}, \mathbf{c}, \mathbf{d}} \quad & J = \mathbf{d}' \left(\sum_{i=1}^n |\mathbf{x}_i|' |\mathbf{x}_i| \right) \mathbf{d} + \xi (\mathbf{a}'\mathbf{a} + \mathbf{c}'\mathbf{c}) \\ \text{subject to} \quad & \begin{cases} \mathbf{a}'x_i + \mathbf{c}'|x_i| + \mathbf{d}'|x_i| \geq y_i + e_i \\ \mathbf{a}'x_i - \mathbf{c}'|x_i| - \mathbf{d}'|x_i| \leq y_i - e_i \\ \mathbf{a}'x_i + \mathbf{c}'|x_i| \leq y_i + e_i \\ \mathbf{a}'x_i - \mathbf{c}'|x_i| \geq y_i - e_i, \quad i = 1, \dots, n \\ \mathbf{c} \geq 0, \quad \mathbf{d} \geq 0 \end{cases} \end{aligned} \quad (9)$$

where ξ is a significantly small positive number, and so leads the influence of $\xi(\mathbf{a}'\mathbf{a} + \mathbf{c}'\mathbf{c})$ to be negligible.

III. ESTIMATION OF POSSIBILITY AND NECESSITY MODELS BY SUPPORT VECTOR MACHINES

In this section, we propose a new method to evaluate interval linear and nonlinear regression models combining the possibility and necessity estimation formulation with the principle of SVM.

A. Interval Linear Regression

We first propose how to get solutions for interval linear regression models by implementing SVM approach for crisp-input interval-output data. Suppose that we are given training data $\{(\mathbf{x}_1, Y_1), \dots, (\mathbf{x}_N, Y_N)\}$, where $\mathbf{x}_i = (x_{i1}, \dots, x_{im})^T$ is the i -th input vector, $Y_i = (y_i, e_i)$ is the corresponding interval output that consists of a center y_i and a radius e_i . For this data set, the possibility and necessity estimation models are denoted as

$$Y^*(\mathbf{x}_i) = B^* + W_1^* x_{i1} + \dots + W_m^* x_{im} \quad i = 1, \dots, N \quad (10)$$

$$Y_*(\mathbf{x}_i) = B_* + W_{*1} x_{i1} + \dots + W_{*m} x_{im} \quad i = 1, \dots, N \quad (11)$$

where the interval coefficients W_i^* and W_{*i} are denoted as $W_i^* = (w_i^*, c_i^*)$ and $W_{*i} = (w_{*i}, c_{*i})$, respectively. Besides, the interval bias coefficients B^* and B_* are denoted as $B^* = (b^*, d^*)$ and $B_* = (b_*, d_*)$, respectively. The estimated interval $Y^*(\mathbf{x}_i)$ by the possibility model always includes the observed interval Y_i , whereas the estimated interval $Y_*(\mathbf{x}_i)$ by the necessity model should be included in the observed interval Y_i . We can denote the interval coefficients as

$$W_i^* = (w_i, c_i + a_i) \text{ and } W_{*i} = (w_i, c_i) \quad (12)$$

which satisfies the condition $W_{*i} \subseteq W_i^*$, $i = 0, 1, \dots, m$, since c_i and a_i are assumed to be nonnegative. Similarly, we can denote the interval bias coefficients as

$$B^* = (b, d + k) \text{ and } B_* = (b, d) \quad (13)$$

which satisfies the condition $B_* \subseteq B^*$, since d and k are assumed to be nonnegative. Therefore, by interval arithmetic the possibility model $Y^*(\mathbf{x}_i)$ and the necessity model $Y_*(\mathbf{x}_i)$ can be written as

$$Y^*(\mathbf{x}_i) = W_1^* x_{i1} + \dots + W_m^* x_{im} + B^* \\ = \langle \mathbf{w} \cdot \mathbf{x} \rangle + b, \langle \mathbf{c} \cdot |\mathbf{x}| \rangle + d + \langle \mathbf{a} \cdot |\mathbf{x}| \rangle + k \quad (14)$$

and

$$Y_*(\mathbf{x}_i) = W_{*1} x_{i1} + \dots + W_{*m} x_{im} + B_* \\ = \langle \mathbf{w} \cdot \mathbf{x} \rangle + b, \langle \mathbf{c} \cdot |\mathbf{x}| \rangle + d \quad (15)$$

The problem of finding the $\mathbf{w}$, $\mathbf{c}$, $\mathbf{a}$, b , d and k that satisfies the inclusion relation $Y^*(\mathbf{x}_i) \supseteq Y_i \supseteq Y_*(\mathbf{x}_i)$ is equivalent to the following constrained optimization problem:

$$\min_{\mathbf{a}, \mathbf{c}, \mathbf{d}} J = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C_1}{2} (\|\mathbf{c}\|^2 + d) + \frac{C_2}{2} (\|\mathbf{a}\|^2 + k) \\ \text{subject to } \begin{cases} \langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b + \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle + d + \langle \mathbf{a} \cdot |\mathbf{x}_i| \rangle + k \geq y_i + e_i \\ \langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b - \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle - d - \langle \mathbf{a} \cdot |\mathbf{x}_i| \rangle - k \leq y_i - e_i \\ \langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b + \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle + d \leq y_i + e_i \\ \langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b - \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle - d \geq y_i - e_i, i = 1, \dots, n \\ k, d \geq 0 \end{cases} \quad (16)$$

The weight coefficient $C_1, C_2 > 0$ determines the tradeoff between model complexity, which is characterized by $\|\mathbf{w}\|^2$, and the vagueness of the possibility and the necessity models, which is characterized by $(\|\mathbf{c}\|^2 + d)$ and $(\|\mathbf{a}\|^2 + k)$, respectively. We can find the solution of this optimization problem given by Eq. (18) in dual variables by finding the saddle point of the Lagrangian:

$$L = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C_1}{2} (\|\mathbf{c}\|^2 + d) + \frac{C_2}{2} (\|\mathbf{a}\|^2 + k) \\ - \sum_{i=1}^N \alpha_{1i} (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b + \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle + d + \langle \mathbf{a} \cdot |\mathbf{x}_i| \rangle + k - y_i - e_i) \\ - \sum_{i=1}^N \alpha_{2i} (-\langle \mathbf{w} \cdot \mathbf{x}_i \rangle - b + \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle + d + \langle \mathbf{a} \cdot |\mathbf{x}_i| \rangle + k + y_i - e_i) \\ - \sum_{i=1}^N \alpha_{3i} (-\langle \mathbf{w} \cdot \mathbf{x}_i \rangle - b - \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle - d + y_i + e_i) \\ - \sum_{i=1}^N \alpha_{4i} (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b - \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle - d - y_i + e_i) \\ - \mu d - \gamma k \quad (17)$$

where $\alpha_{1i}, \alpha_{2i}, \alpha_{3i}, \alpha_{4i}, \mu$, and γ are nonnegative Lagrange multipliers. Differentiating L with respect to $\mathbf{w}$, $\mathbf{c}$, $\mathbf{a}$, b , d , and k , and setting the result to zero, we obtain:

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^N (\alpha_{1i} - \alpha_{2i} + \alpha_{3i} - \alpha_{4i}) \mathbf{x}_i, \quad (18)$$

$$\frac{\partial L}{\partial \mathbf{c}} = 0 \Rightarrow \mathbf{c} = \frac{1}{C_1} \sum_{i=1}^N (\alpha_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) |\mathbf{x}_i|, \quad (19)$$

$$\frac{\partial L}{\partial \mathbf{a}} = 0 \Rightarrow \mathbf{a} = \frac{1}{C_2} \sum_{i=1}^N (\alpha_{1i} + \alpha_{2i}) |\mathbf{x}_i|, \quad (20)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^N (\alpha_{1i} - \alpha_{2i} - \alpha_{3i} + \alpha_{4i}) = 0, \quad (21)$$

$$\frac{\partial L}{\partial d} = 0 \Rightarrow \sum_{i=1}^N (\alpha_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) = C_1 - \mu \leq C_1, \quad (22)$$

$$\frac{\partial L}{\partial k} = 0 \Rightarrow \sum_{i=1}^N (\alpha_{1i} + \alpha_{2i}) = C_2 - \gamma \leq C_2, \quad (23)$$

Substituting Eqs. (18)-(23) into (17), we obtain the following dual problem

$$\max \frac{-1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_{1i} - \alpha_{2i} - \alpha_{3i} + \alpha_{4i}) (\alpha_{1j} - \alpha_{2j} - \alpha_{3j} + \alpha_{4j}) \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle \\ - \frac{1}{2C_1} \sum_{i=1}^N \sum_{j=1}^N (\alpha_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) (\alpha_{1j} + \alpha_{2j} - \alpha_{3j} - \alpha_{4j}) \langle |\mathbf{x}_i| \cdot |\mathbf{x}_j| \rangle \\ - \frac{1}{2C_2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_{1i} + \alpha_{2i}) (\alpha_{1j} + \alpha_{2j}) \langle |\mathbf{x}_i| \cdot |\mathbf{x}_j| \rangle \\ + \sum_{i=1}^N (\alpha_{1i} - \alpha_{2i} - \alpha_{3i} + \alpha_{4i}) y_i + \sum_{i=1}^N (\alpha_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) e_i. \\ \text{subject to} \quad (24)$$

$$\sum_{i=1}^N (\alpha_{1i} - \alpha_{2i} - \alpha_{3i} + \alpha_{4i}) = 0, \\ \sum_{i=1}^N (\alpha_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) \leq C_1 \\ \sum_{i=1}^N (\alpha_{1i} + \alpha_{2i}) \leq C_2, \\ \alpha_{1i}, \alpha_{2i}, \alpha_{3i}, \alpha_{4i} \geq 0.$$

As for the pair $(\mathbf{w}, \mathbf{c}, \mathbf{a}, b, d, p)$, it is easy to find that

$$\mathbf{w} = \sum_{i=1}^N (\alpha_{1i} - \alpha_{2i} + \alpha_{3i} - \alpha_{4i}) \mathbf{x}_i,$$

$$\mathbf{c} = \frac{1}{C_1} \sum_{i=1}^N (\alpha_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) |\mathbf{x}_i|,$$

$$\text{and } \mathbf{a} = \frac{1}{C_2} \sum_{i=1}^N (\alpha_{1i} + \alpha_{2i}) |\mathbf{x}_i|.$$

While parameters b , d and k can be determined from the Karush-Kuhn-Tucker (KKT) conditions:

$$\alpha_{1i} [\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b + \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle + d + \langle \mathbf{a} \cdot |\mathbf{x}_i| \rangle + p - y_i - e_i] = 0 \quad (25)$$

$$\alpha_{2i} [-\langle \mathbf{w} \cdot \mathbf{x}_i \rangle - b + \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle + d + \langle \mathbf{a} \cdot |\mathbf{x}_i| \rangle + p + y_i - e_i] = 0 \quad (26)$$

$$\alpha_{3i} [-\langle \mathbf{w} \cdot \mathbf{x}_i \rangle - b - \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle + d + y_i + e_i] = 0 \quad (27)$$

$$\alpha_{4i} [\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b - \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle - d - y_i + e_i] = 0 \quad (28)$$

For some $\alpha_{3i}, \alpha_{4i} > 0$, the second factor in Eqs. (27) and (28) has to vanish. Hence, b and d can be computed as follows:

$$b = \frac{1}{2} \left(-\langle \mathbf{w} \cdot \mathbf{x}_i \rangle - \langle \mathbf{w} \cdot \mathbf{x}_j \rangle - \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle + \langle \mathbf{c} \cdot |\mathbf{x}_j| \rangle \right), \quad (29)$$

$$d = \frac{1}{2} \left(-\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + \langle \mathbf{w} \cdot \mathbf{x}_j \rangle - \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle - \langle \mathbf{c} \cdot |\mathbf{x}_j| \rangle \right), \quad (30)$$

for some $\alpha_{3i}, \alpha_{4j} > 0$.

Once bias b and d is computed, we can determine the bias k from the Karush-Kuhn-Tucker (KKT) conditions. For some $\alpha_{1i}, \alpha_{2j} > 0$, the second factor in Eqs. (25) and (26) has to vanish. Hence, bias k can be computed as follows:

$$k = -\langle \mathbf{w} \cdot \mathbf{x}_i \rangle - b - \langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle - d - \langle \mathbf{a} \cdot |\mathbf{x}_i| \rangle + y_i + e_i, \quad (31)$$

$$k = \langle \mathbf{w} \cdot \mathbf{x}_j \rangle + b - \langle \mathbf{c} \cdot |\mathbf{x}_j| \rangle - d - \langle \mathbf{a} \cdot |\mathbf{x}_j| \rangle - y_j + e_j, \quad (32)$$

for some $\alpha_{1i}, \alpha_{2j} > 0$.

Therefore, the possibility model $Y^*(\mathbf{x}_i)$ and the necessity model $Y_*(\mathbf{x}_i)$ constructed by the proposed approach are given as follows:

$$Y^*(\mathbf{x}) = \begin{pmatrix} \sum_{i=1}^N (a_{1i} - \alpha_{2i} - \alpha_{3i} + \alpha_{4i}) \langle \mathbf{x} \cdot \mathbf{x}_i \rangle + b, \\ \sum_{i=1}^N (a_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) \langle |\mathbf{x}| \cdot |\mathbf{x}_i| \rangle + d \\ + \sum_{i=1}^N (a_{1i} + \alpha_{2i}) \langle |\mathbf{x}| \cdot |\mathbf{x}_i| \rangle + k \end{pmatrix} \quad (33)$$

$$Y_*(\mathbf{x}) = \begin{pmatrix} \sum_{i=1}^N (a_{1i} - \alpha_{2i} - \alpha_{3i} + \alpha_{4i}) \langle \mathbf{x} \cdot \mathbf{x}_i \rangle + b, \\ \sum_{i=1}^N (a_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) \langle |\mathbf{x}| \cdot |\mathbf{x}_i| \rangle + d \end{pmatrix} \quad (34)$$

B. Extension to the Interval Nonlinear Regression Model

Next, we will consider how to get solutions for interval nonlinear regression models. The basic idea is that an interval nonlinear regression function could be achieved by simply mapping the input patterns $\mathbf{x}_i$ by $\Phi: R^n \rightarrow F$ into a high-dimensional feature space F , and then applying the interval linear regression algorithm in that space [5,6,12]. A key property of the proposed algorithm is that the only quantities that we need to compute are inner products, of the form $\langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle$ and $\langle |\mathbf{x}_i| \cdot |\mathbf{x}_j| \rangle$. The algorithm would only depend on the data through inner products in F , i.e. on functions of the form $\langle \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \rangle$ and $\langle \Phi(|\mathbf{x}_i|) \cdot \Phi(|\mathbf{x}_j|) \rangle$. The functional form of mapping Φ does not need to be known since it is implicitly defined by the choice of kernel function $k(\mathbf{x}, \mathbf{y}) \equiv \langle \Phi(\mathbf{x}) \cdot \Phi(\mathbf{y}) \rangle$. Therefore, it is enough to know and use $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \rangle$ and $k(|\mathbf{x}_i|, |\mathbf{x}_j|) = \langle \Phi(|\mathbf{x}_i|) \cdot \Phi(|\mathbf{x}_j|) \rangle$

instead of defining $\Phi(\bullet)$ explicitly. The well-used kernels for the regression problem are given as follows:

$k(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x} \cdot \mathbf{y} \rangle + 1)^d$: Polynomial kernel of degree d ;

$k(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right)$: Gaussian RBF kernel;

$k(\mathbf{x}, \mathbf{y}) = \tanh(k\langle \mathbf{x} \cdot \mathbf{y} \rangle + \theta)$: Sigmoid kernel.

Hence, by replacing $\langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle$ and $\langle |\mathbf{x}_i| \cdot |\mathbf{x}_j| \rangle$ with $k(\mathbf{x}_i, \mathbf{x}_j)$ and $k(|\mathbf{x}_i|, |\mathbf{x}_j|)$, respectively, we obtain the dual quadratic optimization problem given by Eq. (26). Here, we should note that the constraints are not changed.

$$\begin{aligned} \max \quad & \frac{-1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_{1i} - \alpha_{2i} - \alpha_{3i} + \alpha_{4i})(\alpha_{1j} - \alpha_{2j} - \alpha_{3j} + \alpha_{4j}) k(\mathbf{x}_i, \mathbf{x}_j) \\ & - \frac{1}{2C_1} \sum_{i=1}^N \sum_{j=1}^N (\alpha_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i})(\alpha_{1j} + \alpha_{2j} - \alpha_{3j} - \alpha_{4j}) k(|\mathbf{x}_i|, |\mathbf{x}_j|) \\ & - \frac{1}{2C_2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_{1i} + \alpha_{2i})(\alpha_{1j} + \alpha_{2j}) k(|\mathbf{x}_i|, |\mathbf{x}_j|) \\ & + \sum_{i=1}^N (\alpha_{1i} - \alpha_{2i} - \alpha_{3i} + \alpha_{4i}) y_i + \sum_{i=1}^N (\alpha_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) e_i. \end{aligned} \quad (35)$$

subject to

$$\sum_{i=1}^N (\alpha_{1i} - \alpha_{2i} - \alpha_{3i} + \alpha_{4i}) = 0,$$

$$\sum_{i=1}^N (\alpha_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) \leq C_1$$

$$\sum_{i=1}^N (\alpha_{1i} + \alpha_{2i}) \leq C_2,$$

$$\alpha_{1i}, \alpha_{2i}, \alpha_{3i}, \alpha_{4i} \geq 0.$$

By using different kernel functions $k(\mathbf{x}, \mathbf{y})$, one can construct different learning machines with arbitrary types of nonlinear regression functions in the input space [5,6,12]. Therefore, the possibility model $Y^*(\mathbf{x}_i)$ and the necessity model $Y_*(\mathbf{x}_i)$ constructed by the proposed approach are given as follows:

$$Y^*(\mathbf{x}) = \begin{pmatrix} \sum_{i=1}^N (a_{1i} - \alpha_{2i} - \alpha_{3i} + \alpha_{4i}) k(\mathbf{x}, \mathbf{x}_i) + b, \\ \sum_{i=1}^N (a_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) k(|\mathbf{x}|, |\mathbf{x}_i|) + d \\ + \sum_{i=1}^N (a_{1i} + \alpha_{2i}) k(|\mathbf{x}|, |\mathbf{x}_i|) + k \end{pmatrix} \quad (36)$$

and

$$Y_*(\mathbf{x}) = \begin{pmatrix} \sum_{i=1}^N (a_{1i} - \alpha_{2i} - \alpha_{3i} + \alpha_{4i}) k(\mathbf{x}, \mathbf{x}_i) + b, \\ \sum_{i=1}^N (a_{1i} + \alpha_{2i} - \alpha_{3i} - \alpha_{4i}) k(|\mathbf{x}|, |\mathbf{x}_i|) + d \end{pmatrix} \quad (37)$$

Kernel representations offer an alternative solution by projecting the data into a high dimensional feature space to increase the computational power of the interval linear regression algorithm. The use of interval linear regression model in the dual representation makes it possible to perform this step implicitly. By replacing the inner product with an appropriately chosen “kernel” function, one can implicitly perform a nonlinear mapping to a high dimensional feature space without increasing the number of tunable parameters [13,14]. Another attraction of the kernel method is that the learning algorithms can largely decoupled from the specifics of the application area, which must simply be encoded into the design of an appropriate kernel function. Hence, the problem of choosing an architecture for an interval regression model is replaced by the problem of choosing a suitable kernel. Several kernels have been developed by other researchers for discrete structures such as text and bio-sequences [15,16], indicating that the proposed approach is not restricted only to input spaces that are subsets of the Euclidean space.

IV. NUMERICAL EXAMPLE

In this section, we use one data set, crisp inputs-interval outputs shown in Table I. This data is taken from Tanaka and Lee [4]. The proposed SVM approaches are applied to this data set from an illustrative point of view.

TABLE I. CRISP INPUT-INTERVAL OUTPUT DATA

No.	1	2	3	4	5	6	7	8
x	1	2	3	4	5	6	7	8
Y	[15,30]	[20,37.5]	[15,35]	[25,60]	[25,55]	[40,65]	[55,95]	[70,100]

We first consider interval linear regression analysis. The interval linear model for this data is assumed as

$$Y(\mathbf{x}) = B + W \cdot \mathbf{x}$$

Applying the proposed SVM approach to the given data set, we obtain the possibility model $Y^*(\mathbf{x})$ and the necessity model $Y_*(\mathbf{x})$ as

$$Y^*(\mathbf{x}) = (10.7813, 9.5833) + (7.9167, 29.1667)x$$

$$Y_*(\mathbf{x}) = (10.7813, 0.6250) + (7.9167, 3.6458)x$$

Here, we set $C_1=700$ and $C_2=400$, respectively. We now introduce Tanaka and Lee’s results to compare and evaluate both methods for the linear case

$$Y^*(\mathbf{x}) = (7.3111, 11.856) + (8.371, 2.462)x$$

$$Y_*(\mathbf{x}) = (7.311, 0.168) + (8.371, 0.514)x$$

We can see the interval coefficients are similar. A detailed comparison will be made later in Table II by using a measure of fitness. Fig. 1 illustrates the results for the proposed linear model. The solid line explains the fitted regression line for center. The two dash and dotted lines explain the fitted possibility and necessity models, respectively. As seen from Fig. 1, the proposed interval linear regression method works fairly well.

We now apply the proposed nonlinear method to analyze this data set. At first glance over Fig. 1, the nonlinear regression model looks appropriate for these data. Hence, we apply interval nonlinear regression method to this data set. Here, we use a Gaussian kernel for estimating the interval nonlinear model. Here, we set $C_1=1$, $C_2=100000$, and $\sigma=0.08$, respectively. Fig. 2 illustrates the results for the proposed nonlinear model.

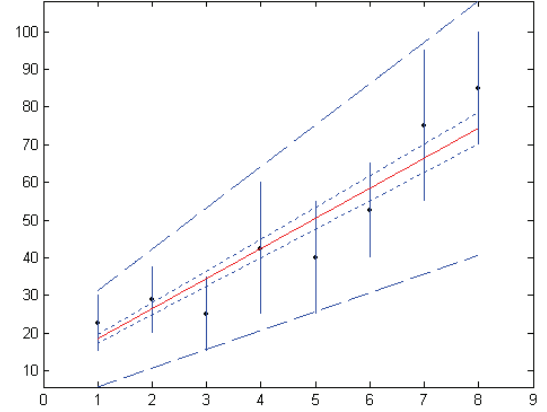


Figure 1. Linear interval regression model obtained by the proposed approach.

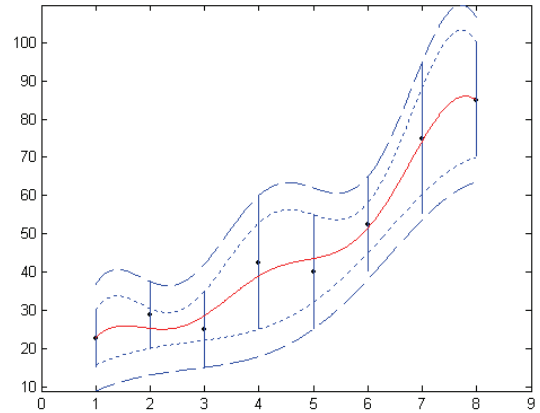


Figure 2. Linear interval regression model obtained by the proposed approach with Gaussian kernel.

As in Fig. 1, the solid line explains the fitted regression curve for center. The two dash and dotted curves explain the fitted possibility and necessity models, respectively. As seen from Fig. 2, we know that the interval nonlinear model looks more appropriate for this particular data set, and the proposed nonlinear algorithm performs well, too.

We now consider a measure of fitness to evaluate both the proposed methods and the methods of Tanaka and Lee’s. We briefly review the measure of fitness (for details, see [4]). In the case of the linear model, the measure of fitness for the i -th data $\phi_Y(\mathbf{x}_i)$ is defined as

$$\varphi_Y(\mathbf{x}_i) = \frac{\langle \mathbf{c} \cdot \mathbf{x}_i \rangle + d}{\langle \mathbf{c} \cdot \mathbf{x}_i \rangle + d + \langle \mathbf{a} \cdot \mathbf{x}_i \rangle + k}$$

which indicates how closely the possibility output for the i -th input approximates the necessity output for the i -th input. Here, the denominator and the numerator represent the spreads of the estimated intervals $Y^*(\mathbf{x}_i)$ and $Y_*(\mathbf{x}_i)$, respectively. Then, the measure of fitness for all data φ_Y is defined as

$$\varphi_Y = \frac{1}{n} \sum_{i=1}^n \varphi_Y(\mathbf{x}_i) = \frac{1}{n} \sum_{i=1}^n \frac{\langle \mathbf{c} \cdot \mathbf{x}_i \rangle + d}{\langle \mathbf{c} \cdot \mathbf{x}_i \rangle + d + \langle \mathbf{a} \cdot \mathbf{x}_i \rangle + k}$$

Where n is a sample size and $0 \leq \varphi_Y \leq 1$. We can also define $\varphi_Y(\mathbf{x}_i)$ and φ_Y for nonlinear model by using the corresponding spreads of the estimated intervals $Y^*(\mathbf{x}_i)$ and $Y_*(\mathbf{x}_i)$. The larger the value of φ_Y , the better the model is fitting the data.

Table II shows the obtained $\varphi_Y(\mathbf{x}_i)$ and φ_Y for the linear model and the quadratic model of Tanaka and Lee's, and the proposed linear and nonlinear models. We can observe that the values of $\varphi_Y(\mathbf{x}_i)$ in the proposed nonlinear models are much larger than those in the linear and quadratic models of Tanaka and Lee's. Also, the summary measure of fit for all data φ_Y is greater for the nonlinear model as compared to the linear model. In particular, the nonlinear model with Gaussian kernel has the largest value $\varphi_Y = 0.5699$. Thus, the measure of fit indicates that the nonlinear model is providing the best fit to the given data compared to the other models. As a whole, the proposed interval regression and model selection methods work fairly well.

TABLE II. OBTAINED $\varphi_Y(\mathbf{x}_i)$ AND φ_Y FOR NUMERICAL DATA

No.(i)	$\varphi_Y(\mathbf{x}_i)$		
	linear		Gaussian kernel
	Tanaka's approach	Proposed approach	Proposed approach
1	0.0476	0.0900	0.5125
2	0.0713	0.1005	0.4284
3	0.0889	0.1076	0.4750
4	0.1025	0.1128	0.6613
5	0.1133	0.1168	0.6197
6	0.1221	0.1198	0.4895
7	0.1295	0.1223	0.6761
8	0.1356	0.1244	0.6967
φ_Y	0.1013	0.1118	0.5699

V. CONCLUSION

Through numerical experiments, we realize that the proposed algorithms derive the satisfying solutions and are the attractive approaches to modeling the data with crisp inputs and interval output. In particular, we realize that we can use these algorithms successfully when a linear model is inappropriate. One of advantages of these algorithms is that we do not need to assume the underlying structure for the interval nonlinear regression model used in this paper. The algorithms combine generalization control with a technique to

address the curse of dimensionality. The main formulation results in solving a global quadratic optimization problem with linear constraints. Hence, this is not a computationally expensive way. In fact, this quadratic optimization problem can be solved easily by such computer software as Matlab, Mathematica, and so on.

In this paper, we deal with estimating interval nonlinear regression model combining the possibility and necessity estimation formulation with the principle of SVM for crisp inputs and interval output dataset. In future work, we intend to devise algorithms for estimating interval nonlinear regression model with interval inputs and interval output, without assuming the underlying structure. Furthermore, we will devise algorithms for estimating fuzzy nonlinear regression model with fuzzy inputs and fuzzy output.

ACKNOWLEDGMENT

This research work was supported in part by the National Science Council Research Grant NSC 99-2221-E-151 -058 -.

REFERENCES

- [1] H. Tanaka, S. Uejima, and K. Asai, "Linear regression analysis with fuzzy model," IEEE Trans. Syst. Man, Cybern., vol. 12, pp. 903–907, 1982.
- [2] H. Tanaka, "Fuzzy data analysis by possibilistic linear models," Fuzzy Sets Syst., vol. 24, pp. 363–375, 1987.
- [3] H. Tanaka and J. Watada, "Possibilistic linear systems and their application to the linear regression model," Fuzzy Sets Syst., vol. 27, pp. 275–289, 1988.
- [4] H. Tanaka and H. Lee, "Interval regression analysis by quadratic programming approach," IEEE Trans. Fuzzy System. vol. 6, no. 4, pp. 473–481, 1998.
- [5] V. Vapnik, The Nature of Statistical Learning Theory. Berlin, Germany: Springer-Verlag, 1995.
- [6] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," NeuroCOLT, NeuroCOLT2 Tech. Rep., 1998.
- [7] J.-T. Jeng, C.-C. Chuang, and S.-F. Su, "Support vector interval regression networks for interval regression analysis," Fuzzy Sets Syst., vol. 138, pp. 283–300, 2003.
- [8] C. Hwang, D. H. Hong, and K. H. Seok, "Support vector interval regression machine for crisp input and output data," Fuzzy Sets and Systems, vol. 157, pp. 1114–1125, 2006.
- [9] P.-Y. Hao, "Interval regression analysis using support vector networks," Fuzzy Sets and Systems, vol. , pp. , 2009.
- [10] H. Ishibuchi, H. Tanaka, "Several formulations of interval regression analysis," Proc. Sino-Japan Joint Meeting on Fuzzy Sets and Systems, Beijing, China, 1990, pp. B2–2.
- [11] G. Peters, "Fuzzy linear regression with fuzzy intervals," Fuzzy Sets and Systems, vol. 63, pp. 45–54, 1994.
- [12] H. Drucker, C. Burges, L. Kaufman, A. Smola, and V. N. Vapnik, "Support vector regression machines," In Advances in Neural Information Processing Systems 9, vol. 9, pp. 155–161. The MIT Press, 1996.
- [13] B. Schölkopf, C. J. C. Burges, and A. J. Smola, Advances in Kernel Method - Support Vector Learning, MIT Press, Cambridge, MA, 1999.
- [14] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, 2000.
- [15] A. Ben-Hur and W. S. Noble, "Kernel methods for predicting protein-protein interactions," Bioinformatics, vol. 21, suppl. 1, pp. i38–46, 2005.
- [16] H. Saigo, J.-P. Vert, N. Ueda, and T. Akutsu, "Protein homology detection using string alignment kernels," Bioinformatics, vol. 20, no. 11, pp. 1682–1689, 2004.