Fuzzy Support Vector Regression

Yahya Forghani
Computer Department,
Mashhad Branch, Azad University,
Mashhad, Iran
yahyafor2000@yahoo.com

Hadi Sadoghi Yazdi
Computer department
Ferdowsi University of Mashhad
Mashhad, Iran
h-sadoghi@um.ac.ir

Reza Sigari Tabrizi
Faculty of Management
Multimedia University
Malaysia
reza.tabrizi@live.com

Mohammad -R. Akbarzadeh -T
Computer department
Ferdowsi University of Mashhad
Mashhad, Iran
akbarzadeh@ieee.org

***Abstract*—The epsilon-SVR has two limitations. Firstly, the tube radius (epsilon) or noise rate along the y-axis must be already specified. Secondly, this method is suitable for function estimation according to training data in which noise is independent of input x (is constant). To resolving these limitations, in approaches like v-SVIRN, the tube radius or the radius of estimated interval function which can be variable with respect to input x, is determined automatically. Then, for the test sample x, the centre of interval function is reported as the most probable value of output according to training samples. This method is useful when the noise of data along the y-axis has a symmetric distribution. In such situation, the centre of interval function and the most probable value of function are identical. In practice, the noise of data along the y-axis may be from an asymmetric distribution. In this paper, we propose a novel approach which estimates simultaneously an interval function and a triangular fuzzy function. The estimated interval function of our proposed method is similar to the estimated function of v-SVIRN. The center of triangular fuzzy function is the most probable value of function according to training samples which is important when the noise of training data along the y-axis is from an asymmetric distribution.**

Keywords- Fuzzy; Interval; Support vector machines (SVMs); Support vector regression machines.

I. INTRODUCTION

Support vector machines [1], [2], [3] is a learning method used for patterns classification and function estimation. The support vector machines are used in various fields due to its simple structure and satisfactory performance. A version of SVM for regression analysis was initiated by [4], [2], [5]. This method is named ε-insensitive support vector regression or ε-SVR. In this method, for training data $\{(x_i, y_i),\ i = 1,2,\dots,n\}$, the objective function $f(.) = w^T\varphi(x) + b$ is estimated such that for each i, $f(x_i) - \varepsilon \le y_i \le f(x_i) + \varepsilon$, where w is a weight vector, b is a bias, and $\varphi(.)$ is a function that map the input space into a high-dimensional feature space. In other words, this method estimates an interval function or a tube with the center $f(.)$ and the radius ε such that maximum possible amount of data are included. This method is faced with following two issues: (1) the tube radius (ε) or the noise rate along the y-axis must be already specified. (2) The method is suitable for those training data in which noise is independent of input value x because the radius of tube is considered to be constant along the x-axis.

Several researchers have made great efforts to solve the mentioned problems of ε-SVR. For example, in support vector interval regression machine or SVIRNs, [6] the lower and upper bound of the tube is determined automatically using two independent RBF networks. By using this method, there is no need to have a priori knowledge about the noise rate along the y-axis. In addition, in this method, the tube size along x-axis can be variable which is suitable for situation that the noise along the y-axis is dependent on input value x. The initial structure of these two RBF networks is formed using ε-SVR approach, and the back propagation method is employed for adjusting the RBF networks. Another method to solve the problems of ε-SVR was proposed by [7]. The method is called support vector interval regression machine or SVIRM. In this method, the upper and lower bound of the tube are gotten simultaneously based on combining the possibility estimation formulation integrating the property of central tendency with the ε-SVR approach. The proposed approach is robust against outliers' impact on the resulting interval regression models. The SVIRM is theoretically simpler than SVIRNs. Finally, the v-SVIRN method was proposed by [8]. This approach is faster and simpler than two previous introduced methods.

In each of the four mentioned methods, for the test sample x, the centre of tube or estimated interval function is reported as the most probable value of output according to training samples. This method is useful when the noise along the y-axis is from a symmetric distribution (e.g. Uniform or Gaussian distribution). In such situation, the centre of function and the most probable value of function are identical. In practice, the noise of data along the y-axis may be from an asymmetric distribution. In this paper, we propose a novel approach which estimates simultaneously an interval function and a triangular fuzzy function. The estimated interval function of our proposed method is similar to the estimated function of v-SVIRN. The center of triangular fuzzy function is the most probable value of function according to training samples which is important when the noise of training data along the y-axis is from an asymmetric distribution.

II. DEFINITIONS

A. Interval value

Definition 1- interval value is a normalized and continuous fuzzy set whose elements have equal degrees of membership [9]. Let $\tilde{x} = [a, b]$ be an interval value. This interval value can be represented by (m, c) in which $m = \frac{a+b}{2}$ is center of interval and $c = \frac{b-a}{2}$ is radius of interval.

B. Basic operators on intervals

Theorem 1- Let $\tilde{x} = (m, c)$ and $\tilde{y} = (n, d)$ be two interval value and s be a scalar [9]. Based on extension principle [10], we have:

$$\tilde{x} + \tilde{y} = (m + n, c + d),$$
$$\tilde{x} - \tilde{y} = (m - n, c + d),$$
$$s\tilde{x} = (sm, |s|c).$$

C. Triangular fuzzy number

Definition 2- The fuzzy number $\tilde{x}$ with the following membership function

$$\mu_{\tilde{x}}(x) = \begin{cases} \dfrac{x-l}{i-l} & l < x \le i, \\ \dfrac{r-x}{r-i} & i < x < r, \\ 0 & otherwise, \end{cases}$$

is called triangular fuzzy number (TFN) denoted by $(i, l, r)_{TFN}$ in which i, l, and r are called center, left spread, and right spread of TFN [11].

D. Basic operators on Triangular fuzzy numbers

Theorem 2- Let $\tilde{x} = (i, l, r)_{TFN}$ and $\tilde{y} = (j, m, n)_{TFN}$ be two TFNs and s be a scalar. Based on extension principle [10], we have:

$$\tilde{x} + \tilde{y} = (i + j, l + m, r + n)_{TFN},$$
$$\tilde{x} - \tilde{y} = (i - j, l + n, r + m)_{TFN},$$
$$s\tilde{x} = \begin{cases} (si, sl, sr)_{TFN} & s \ge 0, \\ (si, sr, sl)_{TFN} & s < 0. \end{cases}$$

III. v-SVIRN APPROACH

In this model, the objective is to find two functions $f(.)$ and $g(.)$ or interval function $\tilde{f}(x) = [f(x) - g(x), f(x) + g(x)] = (f(x), g(x))$, so that following condition is met:

$$f(x_i) - g(x_i) \le y_i \le f(x_i) + g(x_i), \quad i = 1,2, \dots, n.$$

To obtain this objective, function $\tilde{f}(x)$ is defined as follows:

$$\tilde{f}(x) = \tilde{w}^T \varphi(x) + \tilde{b},$$

where $\tilde{w} = (\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_m)^T$ and $\tilde{b}$ are weight vector and bias term, respectively. Here, $\tilde{w}_i = (w_i, c_i)$ and $\tilde{b} = (b, d)$ are considered to be interval variables where w_i is the center and c_i is the radius of $\tilde{w}_i$, and b is the center and d is the radius of $\tilde{b}$. Based on interval arithmetic, the $\tilde{f}(x)$ can be expressed as $\tilde{f}(x) = (w^T \varphi(x) + b, c^T \varphi(|x|) + d)$ where $w = (w_1, w_2, \dots, w_m)^T$, $= (c_1, c_2, \dots, c_m)^T$, and $|x| = (|x_1|, |x_2|, \dots, |x_m|)^T$. In other words, $f(x) = w^T \varphi(x) + b$ and $g(x) = c^T \varphi(|x|) + d$. Therefore, the $\tilde{f}(x)$ can be written as follows:

$$\tilde{f}(x) = [w^T \varphi(x) + b - (c^T \varphi(|x|) + d), w^T \varphi(x) + b + (c^T \varphi(|x|) + d)].$$

Consequently, similar to ε-SVR model, the following model is defined as:

$$\min_{w,b,\xi,\hat{\xi}} \frac{1}{2}\|w\|^2 + C\left(v\left(\frac{1}{2}\|c\|^2 + d\right) + \frac{1}{n}\sum_{i=1}^{n}(\xi_i + \hat{\xi}_i)\right)$$

$$\text{subject to} \begin{cases} y_i - \{(w^T \varphi(x_i) + b) + (c^T \varphi(|x_i|) + d)\} \le \xi_i, \\ \{(w^T \varphi(x_i) + b) - (c^T \varphi(|x_i|) + d)\} - y_i \le \hat{\xi}_i, \\ \xi_i, \hat{\xi}_i \ge 0, \quad i = 1,2, \dots, n; \end{cases} \tag{1}$$

where $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$, $\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_n)^T$, and C is a penalty term.

The constraints of above problem imply that outliers can lie outside the tube. The penalty term helps to determine margin size and number of samples outside the tube. The v parameter allows at most a v-percent of the data lie outside the tube. The dual of problem (1) can be formulated as follows:

$$\max_{\delta,\hat{\delta}} \sum_{i=1}^{n}(\delta_i - \hat{\delta}_i) y_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\delta_i - \hat{\delta}_i)(\delta_j - \hat{\delta}_j) K(x_i, x_j) - \frac{1}{2Cv}\sum_{i=1}^{n}\sum_{j=1}^{n}(\delta_i + \hat{\delta}_i)(\delta_j + \hat{\delta}_j) K(|x_i|, |x_j|)$$

$$\text{subject to} \begin{cases} \sum_{i=1}^{n}(\delta_i - \hat{\delta}_i) = 0; \\ \sum_{i=1}^{n}(\delta_i + \hat{\delta}_i) = Cv; \\ 0 \le \delta_i \le \dfrac{C}{n}, \quad i = 1,2, \dots, n; \\ 0 \le \hat{\delta}_i \le \dfrac{C}{n}, \quad i = 1,2, \dots, n; \end{cases} \tag{2}$$

where $\delta = (\delta_1, \delta_2, \dots, \delta_n)^T$, $\hat{\delta} = (\hat{\delta}_1, \hat{\delta}_2, \dots, \hat{\delta}_n)^T$, $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)^T$, and $\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_n)^T$ in which $\hat{\gamma}_i$, γ_i, $\hat{\delta}_i$, and δ_i $\forall\ i$ are Lagrange multipliers. According to the optimality conditions of problem (1), for each i and j that $\delta_i, \hat{\delta}_j \in (0, C/n)$, we have:

$$b = -\frac{1}{2}\left(w^T\varphi(x_i) + w^T\varphi(x_j) + c^T\varphi(|x_i|) - c^T\varphi(|x_j|) - y_i - y_j\right), \quad (3)$$

$$d = -\frac{1}{2}\left(w^T\varphi(x_i) - w^T\varphi(x_j) + c^T\varphi(|x_i|) + c^T\varphi(|x_j|) - y_i + y_j\right). \quad (4)$$

Therefore, the lower, center, and upper bounds of the interval function $\tilde{f}(x)$ are as follows, respectively:

$$f(x) - g(x) = \left(\sum_{i=1}^{n}(\delta_i - \hat{\delta}_i)K(x, x_i) + b\right) - \left(\frac{1}{Cv}\sum_{i=1}^{n}(\delta_i + \hat{\delta}_i)K(x, |x_i|) + d\right),$$

$$f(x) = \left(\sum_{i=1}^{n}(\delta_i - \hat{\delta}_i)K(x, x_i) + b\right),$$

$$f(x) + g(x) = \left(\sum_{i=1}^{n}(\delta_i - \hat{\delta}_i)K(x, x_i) + b\right) + \left(\frac{1}{Cv}\sum_{i=1}^{n}(\delta_i + \hat{\delta}_i)K(x, |x_i|) + d\right).$$

In fact, the center and radius of interval function $\tilde{f}(x)$ are identified as $f(x) = \left(\sum_{i=1}^{n}(\delta_i - \hat{\delta}_i)K(x, x_i) + b\right)$ and $g(x) = \left(\frac{1}{Cv}\sum_{i=1}^{n}(\delta_i + \hat{\delta}_i)K(x, |x_i|) + d\right)$, respectively.

IV. Proposed Model

The most probable value of output according to training samples is equal to the center of interval function (which has been estimated by v-SVIRN), when the noise along the y-axis is from a symmetric distribution (e.g. Uniform or Gaussian distribution). However, the noise along the y-axis may be from an asymmetric distribution. In such situation, the center of interval function is not equal to the most probable value of output. The paper then proposes a novel approach to find both the center of interval function and the most probable value of output. According to our approach, the upper, center, and lower bound of the interval function and a triangular fuzzy function are obtained, simultaneously. The support of triangular fuzzy function is equal to the support of interval function. Beside, the center of triangular fuzzy function shows the most probable value of interval output according to training data. In this method, the objective is to find three functions i.e. $f(x)$, $g(x)$, and $h(x)$ such that $f(x_i) - g(x_i) \le y_i \le f(x_i) + h(x_i),\ i = 1,2,\dots,n$. To achieve this goal, function $\tilde{f}(x)$ is defined as follows:

$$\tilde{f}(x) = \tilde{w}^T\varphi(x) + \tilde{b}$$

where $\tilde{w} = (\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_m)^T$ is weight vector and $\tilde{b}$ is bias rate. Meanwhile, $\tilde{w}_i = (w_i, l_i, r_i)_{TFN}$ and $\tilde{b} = (b, d, e)_{TFN}$ are considered as triangular fuzzy numbers where w_i is center, l_i is left spread, and r_i is right spread of $\tilde{w}_i$. Similarly, b is center, d is left spread, and e is right spread of $\tilde{b}$. Assuming $x \ge 0$, if we use kernel functions like Gaussian kernel function, we have $\varphi(x) \ge 0$ [12]. Hence, $\tilde{f}(x) = (w^T\varphi(x) + b, l^T\varphi(x) + d, r^T\varphi(x) + e)_{TFN}$ where $w = (w_1, w_2, \dots, w_m)^T$, $l = (l_1, l_2, \dots, l_m)^T$, and $r = (r_1, r_2, \dots, r_m)^T$. In other words, $f(x) = w^T\varphi(x) + b$, $g(x) = l^T\varphi(x) + d$, and $h(x) = r^T\varphi(x) + e$. Therefore, similar to previous model, the following model is defined:

$$\min_{w,b,\xi}\frac{1}{2}\|w\|^2 + C\left(v\left(\frac{1}{2}\|l\|^2 + d\right) + v\left(\frac{1}{2}\|r\|^2 + e\right) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{4}\xi_{ij}\right)$$

$$\text{subject to}\begin{cases} y_i - \{(w^T\varphi(x_i) + b) + (r^T\varphi(x_i) + d)\} \le \xi_{i1}, \\ \{(w^T\varphi(x_i) + b) - (l^T\varphi(x_i) + e)\} - y_i \le \xi_{i2}, \\ y_i - (w^T\varphi(x_i) + b) \le \xi_{i3},\ i = 1,2,\dots,n; \\ (w^T\varphi(x_i) + b) - y_i \le \xi_{i4},\ i = 1,2,\dots,n; \\ \xi_{ij} \ge 0,\ i = 1,2,\dots,n; j = 1,2,3,4; \end{cases} \quad (5)$$

The Lagrange function of the problem (5) is as follows:

$$L(w,b,l,d,r,e,\xi,\delta,\gamma) = \frac{1}{2}\|w\|^2 + C\left(v\left(\frac{1}{2}\|l\|^2 + d\right) + v\left(\frac{1}{2}\|r\|^2 + e\right) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_{ij}\right) + \sum_{i=1}^{n}\delta_{i1}\left(y_i - (w^T\varphi(x_i) + b) - (r^T\varphi(x_i) + d) - \xi_{i1}\right) + \sum_{i=1}^{n}\delta_{i2}\left((w^T\varphi(x_i) + b) - (l^T\varphi(x_i) + e) - y_i - \xi_{i2}\right) + \sum_{i=1}^{n}\delta_{i3}\left(y_i - (w^T\varphi(x_i) + b) - \xi_{i3}\right) + \sum_{i=1}^{n}\delta_{i4}\left((w^T\varphi(x_i) + b) - y_i - \xi_{i4}\right) - \sum_{i=1}^{n}\sum_{j=1}^{4}\gamma_{ij}\xi_{ij}, \quad (6)$$

where γ_{ij} and $\delta_{ij}\ \forall\ i, j$ are the Lagrange multipliers. From the optimality conditions of program (5), we have:

$$\frac{\partial L(w,b,l,d,r,e,\xi,\delta,\gamma)}{\partial w} = 0 \to w = \sum_{i=1}^{n}(\delta_{i1} - \delta_{i2} + \delta_{i3} - \delta_{i4})\varphi(x_i); \quad (7)$$

$$\frac{\partial L(w,b,l,d,r,e,\xi,\delta,\gamma)}{\partial r} = 0 \to r = \frac{1}{Cv}\sum_{i=1}^{n}\delta_{i1}\varphi(x_i); \quad (8)$$

$$\frac{\partial L(w,b,l,d,r,e,\xi,\delta,\gamma)}{\partial l} = 0 \to l = \frac{1}{Cv}\sum_{i=1}^{n}\delta_{i2}\varphi(x_i); \quad (9)$$

$$\frac{\partial L(w,b,l,d,r,e,\xi,\delta,\gamma)}{\partial b} = 0 \to \sum_{i=1}^{n}(\delta_{i1} - \delta_{i2} + \delta_{i3} - \delta_{i4}) = 0; \quad (10)$$

$$\frac{\partial L(w,b,l,d,r,e,\xi,\delta,\gamma)}{\partial d} = 0 \to \sum_{i=1}^{n}\delta_{i1} = Cv; \quad (11)$$

$$\frac{\partial L(w,b,l,d,r,e,\xi,\delta,\gamma)}{\partial e} = 0 \to \sum_{i=1}^{n}\delta_{i2} = Cv; \quad (12)$$

$$\frac{\partial L(w,b,l,d,r,e,\xi,\delta,\gamma)}{\partial \xi_{ij}} = 0 \to \delta_{ij} = \frac{C}{n} - \gamma_{ij},\ i = 1,2,\dots,n; j = 1,2,3,4; \quad (13)$$

$$y_i - \{(w^T\varphi(x_i) + b) + (r^T\varphi(x_i) + d)\} \le \xi_{i1},\ i = 1,2,\dots,n; \quad (14)$$

$$\{(w^T\varphi(x_i) + b) - (l^T\varphi(x_i) + e)\} - y_i \le \xi_{i2},\ i = 1,2,,\dots,n; \quad (15)$$

$$y_i - (w^T\varphi(x_i) + b) \le \xi_{i3},\ i = 1,2,\dots,n; \quad (16)$$

$$(w^T\varphi(x_i) + b) - y_i \le \xi_{i4},\ i = 1,2,\dots,n; \quad (17)$$

$$\delta_{i1}\left(y_i - (w^T\varphi(x_i) + b) - (r^T\varphi(x_i) + d) - \xi_{i1}\right) = 0,\ i = 1,2,\dots,n; \quad (18)$$

$$\delta_{i2}\left((w^T\varphi(x_i) + b) - (l^T\varphi(x_i) + e) - y_i - \xi_{i2}\right) = 0,\ i = 1,2,\dots,n; \quad (19)$$

$$\delta_{i3}\left(y_i - (w^T\varphi(x_i) + b) - \xi_{i3}\right) = 0,\ i = 1,2,\dots,n; \quad (20)$$

$$\delta_{i4}\left((w^T\varphi(x_i) + b) - y_i - \xi_{i4}\right) = 0,\ i = 1,2,\dots,n; \quad (21)$$

$$\gamma_{ij}\xi_{ij} = 0,\ i = 1,2,\dots,n; j = 1,2,3,4; \quad (22)$$

$$\xi_{ij}, \delta_{ij}, \gamma_{ij} \geq 0,\ i = 1,2, \dots, n; j = 1,2,3,4. \quad (23)$$

With respect to (13) and $\gamma_{ij} \geq 0$, we have:

$$0 \leq \delta_{ij} \leq \frac{C}{n}, \quad i = 1,2, \dots, n; j = 1,2,3,4. \quad (24)$$

Substituting (7)-(13), (22), and (24) into the Lagrange function, we obtain the dual of problem (5) as follows:

$$\max_{\delta} \sum_{i=1}^{n} (\delta_{i1} - \delta_{i2} + \delta_{i3} - \delta_{i4}) y_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\delta_{i1} - \delta_{i2} + \delta_{i3} - \delta_{i4}) \begin{pmatrix} \delta_{j1} - \delta_{j2} \\ +\delta_{j3} - \delta_{j4} \end{pmatrix} K(x_i, x_j) - \frac{1}{2Cv} \sum_{i=1}^{n} \sum_{j=1}^{n} (\delta_{i1}\delta_{j1} + \delta_{i2}\delta_{j2}) K(x_i, x_j)$$

$$\text{subject to} \begin{cases} \sum_{i=1}^{n} (\delta_{i1} - \delta_{i2} + \delta_{i3} - \delta_{i4}) = 0\,; \\ \sum_{i=1}^{n} \delta_{i1} = Cv; \\ \sum_{i=1}^{n} \delta_{i2} = Cv; \\ 0 \leq \delta_{ij} \leq \frac{C}{n},\ i = 1, \dots, n; j = 1,2,,3,4; \end{cases}$$

Solving the above quadratic programming problem, we find the optimal value of Lagrange multipliers. With respect to (18), if $\delta_{i1} > 0$ then $y_i - (w^T \varphi(x_i) + b) = (r^T \varphi(x_i) + d) + \xi_{i1}$. Regarding (13), if $\delta_{i1} < \frac{C}{n}$ then $\gamma_{i1} > 0$; therefore, with respect to (22), $\xi_{i1} = 0$.

Hence, if $0 < \delta_{i1} < \frac{C}{n}$ then

$$b + d = y_i - w^T \varphi(x_i) - r^T \varphi(x_i). \quad (25)$$

With respect to (19), if $\delta_{i2} > 0$ then $(w^T \varphi(x_i) + b) - y_i = (l^T \varphi(x_i) + e) + \xi_{i2}$. Furthermore, regarding (13) if $\delta_{i2} < \frac{C}{n}$ then $\gamma_{i2} > 0$; therefore, with respect to (22), $\xi_{i2} = 0$. Hence, if $0 < \delta_{i2} < \frac{C}{n}$ then:

$$b - e = y_i - w^T \varphi(x_i) + l^T \varphi(x_i). \quad (26)$$

With respect to (20), if $\delta_{i3} > 0$ then $y_i - (w^T \varphi(x_i) + b) = \xi_{i3}$. In addition, regarding (13), if $\delta_{i3} < \frac{C}{n}$ then $\gamma_{i3} > 0$; therefore, with respect to (22), $\xi_{i3} = 0$. Hence, if $0 < \delta_{i3} < \frac{C}{n}$ then:

$$b = y_i - w^T \varphi(x_i). \quad (27)$$

Similarly, if $0 < \delta_{i4} < \frac{C}{n}$ then:

$$b = w^T \varphi(x_i) - y_i. \quad (28)$$

Hence, the upper bound of triangular fuzzy function $\tilde{f}(x)$ is equal to:

$$(w^T \varphi(x_i) + b) + (r^T \varphi(x_i) + d) = \left(\sum_{i=1}^{n} (\delta_{i1} - \delta_{i2} + \delta_{i3} - \delta_{i4}) K(x_i, x) + b\right) + \left(\frac{1}{Cv} \sum_{i=1}^{n} \delta_{i3} K(x, x_i) + d\right), \quad (29)$$

where amount of $b + d$ is obtained using (25). The lower bound of triangular fuzzy function $\tilde{f}(x)$ is equal to:

$$(w^T \varphi(x_i) + b) - (l^T \varphi(x_i) + e) = \left(\sum_{i=1}^{n} (\delta_{i1} - \delta_{i2} + \delta_{i3} - \delta_{i4}) K(x_i, x) + b\right) - \left(\frac{1}{Cv} \sum_{i=1}^{n} \delta_{i4} K(x, x_i) + e\right), \quad (30)$$

where amount of $b - e$ is obtained using (26). The center of triangular fuzzy function $\tilde{f}(x)$ is equal to:

$$(w^T \varphi(x_i) + b) + \frac{(r^T \varphi(x_i) + d) - (l^T \varphi(x_i) + e)}{2}, \quad (31)$$

and the most probable function (peak of triangular fuzzy function) is equal to:

$$w^T \varphi(x_i) + b = \sum_{i=1}^{n} (\delta_{i1} - \delta_{i2} + \delta_{i3} - \delta_{i4}) K(x_i, x) + b, \quad (32)$$

where b can be calculated by (27) or (28). In fact, it can be stated that we obtained an interval function and a triangular fuzzy function. The support of triangular fuzzy function is equal to the support of interval function. Eq. (31) shows peak of triangular fuzzy function whereas (32) represents the center of interval function.

V. EXPERIMENT

In this section, we use the training data as used in [6], [7], [8] to verify the performance of our novel approach. The training data are generated by following equations:

$$y_k = 0.2 \sin(2\pi x_k) + 0.2 x_k^2 + 0.3 + (0.1 x_k^2 + 0.05) e_k,$$

$$x_k = 0.02(k - 1),\ k = 1,2, \dots, 21,$$

where noise e_k is a real number randomly generated in the interval $[-1,1]$. For our experiment, the parameters v and C were chosen 0.3 and 100 respectively. Two parameter values except σ are similar for both models based on v-SVIRN and our novel approach. Figure 1 shows the results of the experiment. From Figure 1 we recognize that our proposed method successfully could find not only the upper, center, and lower bounds of interval function but also the most probable output according to training data.

VI. CONCLUSION

In this paper, we proposed a novel approach which estimates simultaneously an interval function and a triangular fuzzy function. Previous works estimate only an interval function. The estimated interval function of our proposed method is similar to the estimated interval function of recently proposed work v-SVIRN. In v-SVIRN, for the test sample x, the centre of interval function is reported as the most probable value of output according to training samples. This method is useful when the noise of data along the y-axis has a symmetric distribution. In such situation, the centre of interval function and the most probable value of function are identical. The center of triangular fuzzy function obtained by using our proposed method is the most probable value of function according to training samples which is important for us when the noise of training data along the y-axis is from an asymmetric distribution.

REFERENCES

[1] Vapnik, V. N., “The nature of statistical learning theory”. New York, Springer-Verilog, 1995.

[2] Vapnik, V., S. E. Golowich, et al. , “Support vector method for function approximation, regression estimation, and signal processing.” Advances in Neural Information Processing Systems, 1996, 9: 281–287.

[3] Vapnik, V. N., “Statistical learning theory”. New York, John Wiley and Sons, 1998.

[4] Drucker, H., C. J. C. Burges, et al., “Support vector regression machines”. Advances in Neural Information Processing Systems, MIT Press, 1996.

[5] Smola, A. J. and B. Schölkopf., “A tutorial on support vector regression.” Statistics and Computing, 2004, 14(3): 199-222.

[6] Jeng, J.-T., C.-C. Chuang, et al. ,”Support vector interval regression networks for interval regression analysis.” Fuzzy Sets Syst., 2003, 138(2): 283-300.

[7] Hwang, C., D. H. Hong, et al., “Support vector interval regression machine for crisp input and output data.” Fuzzy Sets and Systems , 2006,157: 1114-1125.

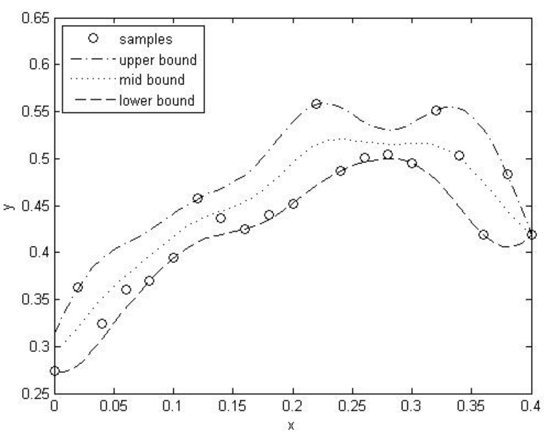
[8] Hao, P.-Y. ,”Interval regression analysis using support vector networks” Fuzzy Sets and Systems, 2009, 160: 2466-2485.

[9] Moore, R. E., R. B. Kearfott, et al., “introduction to interval analysis”. Philadelphia, Siam, 2009.

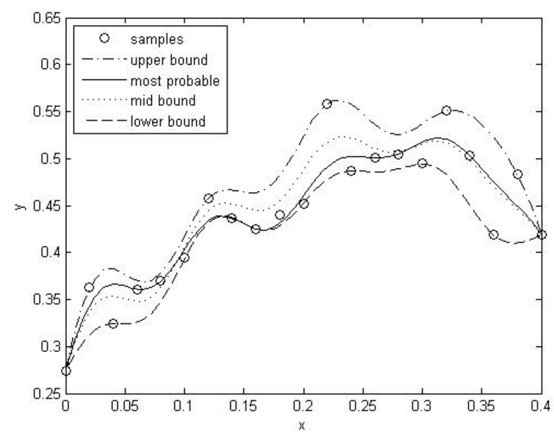
[10] Zadeh, L. A. ,”Fuzzy sets as a basis for a theory of possibility.” Fuzzy Sets and Systems, 1978, 100(Supplement 1): 9-34.

[11] Zimmermann, H.-J. ,”Fuzzy set theory and its applications”, Kluwer Academic Publishers, 1996.

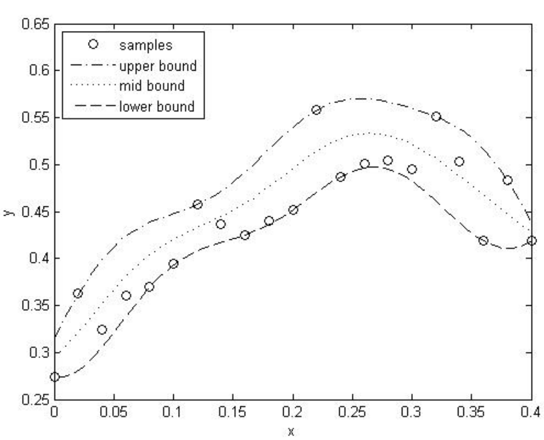
[12] Herbrich, R. ,”Learning Kernel Classifiers: Theory and Algorithms”, MIT Press, Cambridge, 2002.



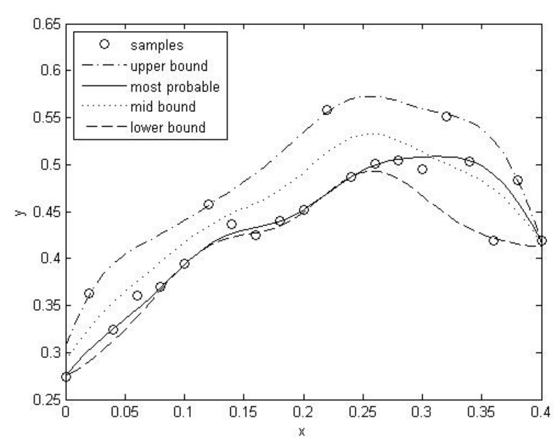
(B1)



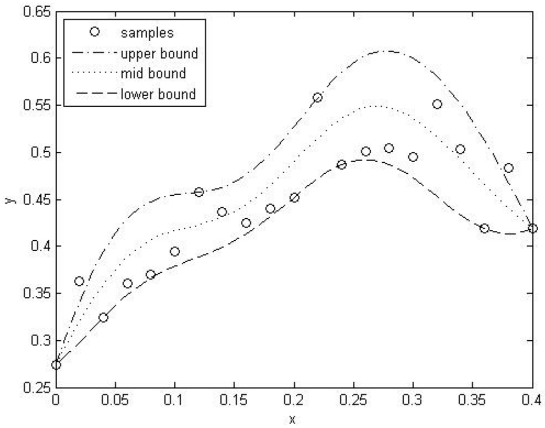
(A1)



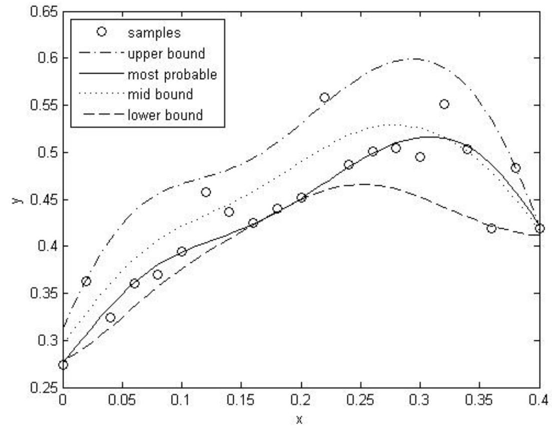
(B2)



(A2)



(B3)



(A3)

Figure 1. Function estimation by using (A) our proposed method and (B) ν-SVIRN, for $\sigma = 0.05$, $\sigma = 0.075$, and $\sigma = 0.1$ respectively.