Decision Making by a Fuzzy Regression Model with Modified Kernel

Kiyoshi Nagata

Faculty of Business Administration
Daito Bunka University
Tokyo 175-8571, Japan
Email: nagata@ic.daito.ac.jp

Michihiro Amagasa

Faculty of Business Administration
Hokkai Gakuen University
Sapporo 062-8605, Japan
Email: amagasa@ba.hokkai-s-u.ac.jp

Abstract—**Regression model is a popular and powerful model for finding a rule from large amount of collected data. It is widely used in various areas for predicting the value derived from observable values. Especially in multivariate numerical analysis, several types of regression models, not only linear but also polynomial or exponential, are established. In case of non-numerical data, although fuzzy regression models are proposed and investigated by some researchers, most of them are linear models. In order to construct a non-linear regression model with fuzzy type data set, new type of devices are needed since fuzzy numbers have a complicated behavior in multiplication and division. In this paper, we try to extend a linear fuzzy regression model to non-linear model by adapting a modified kernel method.**

Keywords–Fuzzy regression model; Kernel method; Decision making.

I. INTRODUCTION

As an analysis method of numerical big-data mining, the regression model is still playing an important role. However, the huge amount of data processing requires strong computing power and resources. In particular, when handling data with non-linear features, finding a proper regression model is not easy, sometimes even infeasible. The kernel method, so-called a kernel trick, is one of smart devices solving this kind of problem. A kernel defined on the product of a data set induces an element of Hilbert space, a space of functions with an inner product, and considering a linear model in the space gives us a non-linear model in the original space. Thus, only the calculation of kernels for the given data set is non-linear, and the calculation for solving the problem to give a model is performed in the linear operation method. The kernel method is applied to many analytical systems, such as the Principal Component Analysis (PCA), [16], the Canonical Correction Analysis (CCA), [6], [12], Fisher's Linear Discriminant Analysis (LDA), [13], the Support Vector Machine(SVM), [1], [7], the regression model, [14], [17], etc.

In the real world, the collected data are sometimes expressed in linguistic values, and in order to apply well-known and authorized stochastic methods such as regression analysis, these values are transformed into numerical data. For instance, the price of a production or a service are determined from several factors, such as price of raw materials, selling expenses, consumer demand, etc. Also the price has high correlations with the customer value of product or service. Bradley T. Gale proposed a scenario where price satisfaction carries 40% of the weight and non-price attributes 60% in the customer-value equation, and showed a figure representing the relationship between relative performance overall score and relative price for luxury cars based on data [9, pp. 218-219]. In that figure, the relative price is generically expressed in linguistic values such as "Higher", "Lower", etc., then these values are transformed into numerical values in order to plot corresponding points on the performance-price plane. For the price prediction model, Inoue et al. proposed a sale price prediction model by fuzzy regression, [11], and Michihiro Amagasa, also proposed a method to handle data with uncertainty in the model of regression analysis as an extension of their model, [3]. We also give a precise formulation of a multi-variable regression model where both explanatory variables and objective one are L-R type fuzzy numbers, [4].

Construction approaches for regression models handling fuzzy set are roughly divided into two types, one is Fuzzy Least Square Regression (FLSR) and the other is dual model for possibilistic regression. The concept of FLSR model is similar to that of ordinary regression model where each value of three vertexes is processed to minimize the sum of distances between the given data and the estimated values. D'Urso adopts this approach handling linear regression model with several types of input-output data, such as crisp-fuzzy, fuzzy-crisp, and fuzzy-fuzzy, with not only type1 fuzzy data set but also type2 fuzzy data set, [8]. The dual model of possibilistic regression approach, originally proposed by Tanaka et al., [18], [19], gives upper and lower regression model by using linear programing analysis approach. Although their model is extended to non-linear models, [10], explanatory variables are still crisp values. In this paper, we propose a non-linear regression model of fuzzy input-fuzzy output type as an extension of our previously proposed model in [4] by applying the kernel method.

The rest of the paper is organized as follows: In Section II, we will review general theory of the kernel method and give a concrete construction of quadratic kernel for a small number of variables. Section III is dedicated to a brief explanation of Guo and Tanaka's non-linear fuzzy regression model and the details of our linear model. Then, in Section IV, we describe the extension version of our model into non-linear type with modified kernels. Examples to see how the proposed model works are coming up with some discussions. The last section, Section V, is the conclusion and the future works.

II. KERNEL THEORY

First, we give a brief description of kernel theory, then give an expression of the functions in the reproducing kernel Hilbert space for a quadratic kernel.

A. Overview of Kernel Theory

For any set $\mathcal{X}$ and the Hilbert space $\mathcal{H}$ of functions on $\mathcal{X}$ over $\mathbb{R}$, a positive definite kernel is a map

$$k : \mathcal{X} \times \mathcal{X} \to \mathcal{H}$$

satisfying

- $k(x, y) = k(y, x)$ for any $x,\ y \in \mathcal{X}$,
- For any $\{c_i\} \subset \mathbb{R}$ and any $\{x_i\} \subset \mathcal{X}$,

$$\sum c_i c_j k(x_i, x_j) \geq 0.$$

Here, we give some examples of kernel over $\mathbb{R}^k$.
For $\vec{x} = (x_1, \ldots, x_k), \vec{y} = (y_1, \ldots, y_k)$,

- $k(\vec{x}, \vec{y}) = \vec{x}^t\vec{y} = \sum_{i=1}^k x_i y_i$ (linear kernel)
- $k_P(\vec{x}, \vec{y}) = (\vec{x}^t\vec{y} + c)^d$,
  with $c \geq 0,\ 0 < d \in \mathbb{Z}$ (polynomial kernel)
- $k_E(\vec{x}, \vec{y}) = exp(\beta\vec{x}^t\vec{y})$, with $\beta > 0$ (exponential kernel)
- $k_G(\vec{x}, \vec{y}) = exp(-\frac{1}{2\sigma^2}\|\vec{x} - \vec{y}\|^2)$
  (Gaussian radial basis function kernel)
- $k_L(\vec{x}, \vec{y}) = exp(-\alpha\sum_{i=1}^k |x_i - y_i|)$ (Laplacian kernel)

If, for any $x \in \mathcal{X}$, there exists a function $k_x \in \mathcal{H}$ such that

$$f(x) = < f, k_x >_{\mathcal{H}}, (\forall f \in \mathcal{H}) \qquad (1)$$

where $< \cdot, \cdot >_{\mathcal{H}}$ is the inner product of the Hilbert space, the Hilbert space $\mathcal{H}$ is called a Reproducing Kernel Hilbert Space (RKHS). It is shown that $k_x \in \mathcal{H}$ is unique, and $k(\cdot, x) = k_x$ is a positive definite kernel on $\mathcal{X}$ called the reproducing kernel.

Conversely, the following theorem is known, [5].

Theorem 1. *(Moore-Aronszajn) For any positive definite kernel on $\mathcal{X}$, there exist unique Hilbert space $\mathcal{H}$ satisfying*

1) *$k(\cdot, x) \in \mathcal{H}$ (for any $x \in \mathcal{X}$),*
2) *The subspace spanned by $\{k(\cdot, x); x \in \mathcal{X}\}$ is dense in $\mathcal{H}$,*
3) *k is the reproducing kernel of $\mathcal{H}$.*

Although Hilbert space has infinity dimension, solution of some optimization problem with data, if there is any, can be expressed as a linear combination of at most the number of data elements in $\mathcal{H}$. This is guaranteed by the following theorem, [15].

Theorem 2. *(The Representer Theorem) Let k be a kernel on $\mathcal{X}$ and let $\mathcal{H}$ be its associated RKHS. Fix $x_1, \cdots, x_n \in \mathcal{X}$, and consider the optimization problem*

$$\min_{f \in \mathcal{H}} D(f(x_1), \ldots, f(x_n)) + P(\|f\|^2_{\mathcal{H}}) \qquad (2)$$

where P is nondecreasing and D depends only on $f(x_1), \cdots, f(x_n)$. If there is a minimizer, then it has the form of

$$f = \sum_{i=1}^n a_i k(\cdot, x_i) \qquad (3)$$

with some $a_1, \cdots, a_n \in \mathbb{R}$. Furthermore, if P is strictly increasing, then every solution has this form.

B. Example Expression of RKHS Basis

From the representer theorem, we can express an optimal function as in the form of (3). However, if the given data set is big, we will have many unknown variables $\{a_i\}_{i=1,\ldots,n}$ to be determined. For the convenience of calculation, we try to reduce the number of components for the polynomial kernel and give an example for the quadratic polynomial kernel of the case that $d = 2$ and $k = 3$ variables.

From the representer theorem and the equation below,

$$k_P(\vec{x}, \vec{y}) = (\textstyle\sum_{j=1}^k x_j y_j + c)^d$$
$$= \sum_{\substack{0 \leq e_1 + \cdots + e_k \leq d \\ 0 \leq e_j}} c^{d-(e_1+\cdot+e_k)} x_1^{e_1} \cdots x_k^{e_k} y_1^{e_1} \cdots y_k^{e_k}$$

we have that for any $(e_1, \ldots, e_k)$ such that $0 \leq e_1 + \cdots + e_k \leq d, 0 \leq e_i$, there exist $N =_{k+d} C_d$ vectors, $\vec{x}_1, \ldots \vec{x}_N$, and $a_1, \ldots, a_N$ satisfying

$$\sum_{i=1}^N a_i x_{i1}^{f_1} \cdots x_{ik}^{f_k} = \begin{cases} c^{-(d-(e_1+\cdot+e_k))} & \text{if}(f_1 \ldots, f_k) = (e_1, \ldots, e_k), \\ 0 & \text{otherwise.} \end{cases} \qquad (4)$$

In a simple case of $d = 2$ and $k = 3$ then $N =_5 C_2 = 10$, and the left side of equation (4) is expressed as

$$\begin{pmatrix} x_{11}^2 & x_{21}^2 & \cdots & x_{10\,1}^2 \\ x_{12}^2 & x_{22}^2 & \cdots & x_{10\,2}^2 \\ x_{13}^2 & x_{23}^2 & \cdots & x_{10\,3}^2 \\ x_{11} & x_{21} & \cdots & x_{10\,1} \\ x_{12} & x_{22} & \cdots & x_{10\,2} \\ x_{13} & x_{23} & \cdots & x_{10\,3} \\ x_{11}x_{12} & x_{21}x_{22} & \cdots & x_{10\,1}x_{10\,2} \\ x_{11}x_{13} & x_{21}x_{23} & \cdots & x_{10\,1}x_{10\,3} \\ x_{12}x_{13} & x_{22}x_{23} & \cdots & x_{10\,2}x_{10\,3} \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_{10} \end{pmatrix}.$$

However, we only have to determine $\vec{x}_1, \vec{x}_2, \vec{x}_3$ and solve the 10 equations of (4) shown as follows.

$$\begin{pmatrix} x_{1j}^2 & x_{2j}^2 & x_{3j}^2 \\ x_{1j} & x_{2j} & x_{3j} \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ c \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ c^2 \end{pmatrix},$$

or

$$\begin{pmatrix} x_{1j}^2 & x_{2j}^2 & x_{3j}^2 \\ x_{1j} & x_{2j} & x_{3j} \\ x_{1j}x_{1l} & x_{2j}x_{2l} & x_{3j}x_{3l} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

where $j, l = 1, 2, 3$ and $j \neq l$. Just analyzing the invertibility of these matrices, we have 10 functions spanning the dense subspace $\mathcal{H}'_k$ of $\mathcal{H}_k$.

$$\mathcal{H}'_k = < k(\cdot, \vec{x}_i); i = 1, \ldots, 10 >_{\mathbb{R}},$$

where
$\vec{x}_1 = (1, 0, 0)$, $\vec{x}_2 = (0, 1, 0)$, $\vec{x}_3 = (0, 0, 1)$,
$\vec{x}_4 = (-1, 0, 0)$, $\vec{x}_5 = (0, -1, 0)$, $\vec{x}_6 = (0, 0, -1)$,
$\vec{x}_7 = (1, 1, 0)$, $\vec{x}_8 = (0, 1, 1)$, $\vec{x}_9 = (1, 0, 1)$, and
$\vec{x}_{10} = (0, 0, 0)$.

III. Some Existing Fuzzy Regression Model

In this section, we will give a brief explanation of two fuzzy regression models, one is crisp-input and fuzzy-output type by Guo and Tanaka, and the other is fuzzy-input and fuzzy-output type.

A. Guo and Tanaka's Non-Linear Model

Guo and Tanaka have investigated the dual possibilistic regression models of both linear and non-linear types with crisp-input and symmetric triangular fuzzy-output in [10]. At first, the linear model whose output $Y = (y;p)_F = (y;p,p)_F$ from crisp input values for variables x_j $(j = 1,\ldots,k)$ is defined as follows,

$$Y = A_1x_1 + A_2x_2 + \cdots + A_kx_k, \tag{5}$$

with symmetric fuzzy coefficients $A_j = (a_j;r_j)_F$ $(j = 1,\ldots,k)$. In this formula, the value of Y is obtained by calculating $(\sum_{j=1}^k a_jc_j, \sum_{j=1}^k r_j|c_j|)$, once explicit values $c_1,\ldots,c_k$ for each given variable. When we have a data set of n number of data, $\{(Y_i;x_{i1},\cdots,x_{ik})\}_{i=1,\ldots,n}$ with crisp x_{ij} and symmetric fuzzy numbers $Y_i = (y_i;p_i)_F$, we consider the upper regression model and the lower regression model.

For the upper regression model, try to find fuzzy coefficients $A_j^* = (a_j^*;r_j^*)_F$ such that

$$\begin{array}{l} \text{Minimizing: } J(\vec{r}^*) = \sum_{j=1}^k r_j^*\left(\sum_{i=1}^n |x_{ij}|\right), \\ \text{under the condition} \\ Y_i \subseteq Y_i^* = A_1^*x_{i1} + \cdots + A_k^*x_{ik} \qquad (i = 1,\ldots,n). \end{array} \tag{6}$$

The inclusion condition above can be expressed by the following equations, because the shapes of fuzzy set are supposed to be similar

$$\begin{cases} y_i - p_i \ge \sum_{j=1}^k a_j^*x_{ij} - \sum_{j=1}^k r_j^*|x_{ij}| \\ y_i + p_i \le \sum_{j=1}^k a_j^*x_{ij} + \sum_{j=1}^k r_j^*|x_{ij}| \\ r_j^* \ge 0 \end{cases}. \tag{7}$$

For the lower regression model, try to find fuzzy coefficients $A_{j*} = (a_{j*};r_{j*})_F$ such that

$$\begin{array}{l} \text{Maximizing: } J(\vec{r_*}) = \sum_{j=1}^k r_{j*}\left(\sum_{i=1}^n |x_{ij}|\right), \\ \text{under the condition} \\ Y_i \supseteq Y_{i*} = A_{1*}x_{i1} + \cdots + A_{k*}x_{ik} \qquad (i = 1,\ldots,n). \end{array} \tag{8}$$

The inclusion condition above also can be expressed by the following equations.

$$\begin{cases} y_i - p_i \le \sum_{j=1}^k a_{j*}x_{ij} - \sum_{j=1}^k r_{j*}|x_{ij}| \\ y_i + p_i \ge \sum_{j=1}^k a_{j*}x_{ij} + \sum_{j=1}^k r_{j*}|x_{ij}| \\ r_{j*} \ge 0 \end{cases}. \tag{9}$$

For the existence of upper and lower regression model, Guo and Tanaka showed the following theorem.

Theorem 3. *(by Guo and Tanaka, [10])*

1) *There always exists an optimal solution in the upper regression model (6) under (7) .*
2) *There exists an optimal solution in the lower regression model (8) under (9) if and only if there exist $a_{1*}^{(0)}, a_{2*}^{(0)}, \ldots, a_{k*}^{(0)}$ satisfying*

$$y_i - p_i \le \sum_{j=1}^k a_{j*}^{(0)}x_{ij} \le y_i + p_i \qquad (i = 1,\ldots,n). \tag{10}$$

From the theorem, there might not be any optimal solution for the lower regression model. This problem is caused by the relationship between the number of variables and the number of data. They tried to solve the problems by extending the model into non-linear model which has more formal variables x_ix_j $(i,j = 1,\ldots,k)$ in the following formula.

$$Y = A_0 + \sum_{j=1}^k A_jx_j + \sum_{j,l=1}^k A_{jl}x_jx_l, \tag{11}$$

with symmetric fuzzy coefficients $A_j,\ A_{jl}$ $(j,l = 1,\ldots,k)$. The right hand side has a quadratic part when considering x_i variables, however we need to find A_j and A_{jl} for a given data set which minimize or maximize the value, so this might be solved by LP method.

B. Our Linear Model

As a general type of fuzzy number, we consider L-R fuzzy set with monotone decreasing functions satisfying $L(0) = R(0) = 1$ and $L(1) = R(1) = 0$, and denote a L-R fuzzy set by $Y = (y;p,q)_F$, where y is the value giving the maximum uncertainty, e.g., 1, and p, q are left and right range from y, i.e., $y - p$ and $y + q$ give the uncertainty value 0, [2]. We proposed the following type of possibilistic fuzzy regression model

$$Y = A_1X_1 + A_2X_2 + \cdots + A_kX_k, \tag{12}$$

with L-R fuzzy variables $Y = (y;p,q)_F$ and $X_j = (x_j;w_j,z_j)_F$ and L-R fuzzy coefficients $A_j = (a_j;r_j,s_j)_F$ $(j = 1,\ldots,k)$.

Let $[Y]_h$ be the support of fuzzy number Y above h-cut line, we have

$$\begin{array}{l} [Y]_h = [y - pL^{-1}(h), y + qR^{-1}(h)], \\ [X_j]_h = [x_j - w_jL^{-1}(h), x_j + z_jR^{-1}(h)], \\ [A_j]_h = [a_j - r_jL^{-1}(h), a_j + s_jR^{-1}(h)]. \end{array}$$

Applying commonly known multiplication and summation of L-R fuzzy numbers, we have

$$\begin{array}{rl} [\sum_{j=1}^k A_jX_j]_h = & [\sum_{j=1}^k (a_j - r_jL^{-1}(h))(x_j - w_jL^{-1}(h)), \\ & \sum_{j=1}^k (a_j + s_jR^{-1}(h))(x_j + z_jR^{-1}(h))]_h, \end{array}$$

and the range of the interval, denoted by J, is calculated by subtracting the left end value from the right end value. Then

$$\begin{array}{rcl} J & = & \sum_{j=1}^k \{(z_jR^{-1}(h) + w_jL^{-1}(h))a_j \\ & & +(x_j + z_jR^{-1}(h))R^{-1}(h)s_j \\ & & +(x_j - w_jL^{-1}(h))L^{-1}(h)r_j\}. \end{array}$$

Following Guo and Tanaka, we consider upper and lower models, and describe the inclusion relation of the support of Y_i and that of the obtained fuzzy number in the regression model for a given data set.

Now we let ZW_j, XZ_j, XW_j be as follows,

$$\begin{cases} ZW_j = (\sum_{i=1}^n z_{ij})R^{-1}(h) + (\sum_{i=1}^n w_{ij})L^{-1}(h) \\ XZ_j = ((\sum_{i=1}^n x_{ij}) + (\sum_{i=1}^n z_{ij})R^{-1}(h))R^{-1}(h) \\ XW_j = ((\sum_{i=1}^n x_{ij}) - (\sum_{i=1}^n w_{ij})L^{-1}(h))L^{-1}(h) \end{cases}. \tag{13}$$

Then our upper model Y^* is constructed with $A_j^* = (a_j^*;r_j^*,s_j^*)_F$, such that

$$\begin{array}{l} \text{Minimizing: } J(\mathbb{A}^*) = \sum_{j=1}^k (ZW_ja_j^* + XZ_js_j^* + XW_jr_j^*), \\ \text{where } \mathbb{A}^* = (A_1^*,\ldots,A_k^*), \end{array} \tag{14}$$

 ISBN: 978-1-61208-566-1

under the condition that for all i

$$\begin{cases} y_i - p_i L^{-1}(h) \geq \sum_{j=1}^{k} (a_j^* - r_j^* L^{-1}(h)) \times \\ \qquad (x_{ij} - w_{ij} L^{-1}(h)) \\ y_i + q_i R^{-1}(h) \leq \sum_{j=1}^{k} (a_j^* + s_j^* R^{-1}(h)) \times \\ \qquad (x_{ij} + z_{ij} R^{-1}(h)) \\ r_j^*,\ s_j^* \geq 0 \end{cases} . \tag{15}$$

The lower model Y_* is similarly constructed with $A_{j*} = (a_{j*}; r_{j*}, s_{j*})_F$, such that

$$\text{Maximizing: } J(\mathbb{A}_*) = \sum_{j=1}^{k} (ZW_j a_{j*} + XZ_j s_{j*} + XW_j r_{j*}), \text{ where } \mathbb{A}_* = (A_{1*}, \ldots, A_{k*}), \tag{16}$$

under the condition that for all i

$$\begin{cases} y_i - p_i L^{-1}(h) \leq \sum_{j=1}^{k} (a_{j*} - r_{j*} L^{-1}(h)) \times \\ \qquad (x_{ij} - w_{ij} L^{-1}(h)) \\ y_i + q_i R^{-1}(h) \geq \sum_{j=1}^{k} (a_{j*} + s_{j*} R^{-1}(h)) \times \\ \qquad (x_{ij} + z_{ij} R^{-1}(h)) \\ r_{j*},\ s_{j*} \geq 0 \end{cases} . \tag{17}$$

We could also show the following theorem similar to the Theorem 3 on the existence of models.

Theorem 4. *When $x_{ij} - w_{ij} L^{-1}(h) > 0$ ($i = 1, \ldots, n,\ j = 1, \ldots, k$), then*

1) *There always exists an optimal solution in the upper regression model (14) under (15) .*
2) *There exists an optimal solution in the lower regression model (16) under (17) if and only if there exist $a_{1*}^{(0)}, a_{2*}^{(0)}, \ldots, a_{k*}^{(0)}$ satisfying*

$$\begin{cases} y_i - p_i L^{-1}(h) \leq \sum_{j=1}^{k} (x_{ij} - w_{ij} L^{-1}(h)) a_{j*}^{(0)} \\ y_i + q_i R^{-1}(h) \geq \sum_{j=1}^{k} (x_{ij} + z_{ij} R^{-1}(h)) a_{j*}^{(0)} \end{cases} . \tag{18}$$

Proof.

1) If $x_{ij} - w_{ij} L^{-1}(h) \geq 0$ in (15), then $x_{ij} > 0$ from $w_{ij} \geq 0$ and $0 \leq L^{-1}(h) \leq 1$. Therefore $x_{ij} + z_{ij} R^{-1}(h)$ are also non-negative, and sufficiently large r_j^* and s_j^* satisfy the condition.
2) If there exist $A_{j*} = (a_{j*}; r_{j*}, s_{j*})_F$ ($j = 1, \ldots, k$) satisfying (17), then we have the condition (18). Conversely, for $a_{j*}^{(0)}$ satisfying (18), put $A_{j*}^{(0)} = (a_{j*}^{(0)}; 0, 0)_F$ and they satisfy the condition (17).

□

Remark1: When the data for independent variables are given in linguistic values, they are usually transformed into fuzzy numbers satisfying the condition $x_{ij} - w_{ij} L^{-1}(h) > 0$ ($i = 1, \ldots, n,\ j = 1, \ldots, k$). So, the assumptions in the Theorem 4 are not special condition.

Remark2: The condition (18) means the inclusion relation between Y_i and the resulted fuzzy number Y_{i*} of areas between h-cut horizontal line and the base-line ($h = 0$) of them.

Remark2.1: In case of $h = 1$, $L^{-1}(1) = R^{-1}(1) = 0$ and (18) is reduced to

$$y_i = \sum_{j=1}^{k} x_{ij} a_{j*}^{(0)},$$

which means that the line segment of Y_{i*} is in the area of Y_i.

Remark2.2: In case of $h = 0$, $L^{-1}(0) = R^{-1}(0) = 1$ and (18) is reduced to

$$\begin{cases} y_i - p_i \leq \sum_{j=1}^{k} (x_{ij} - w_{ij}) a_{j*}^{(0)} \leq \sum_{j=1}^{k} x_{ij} a_{j*}^{(0)} \\ y_i + q_i \geq \sum_{j=1}^{k} (x_{ij} + z_{ij}) a_{j*}^{(0)} \geq \sum_{j=1}^{k} x_{ij} a_{j*}^{(0)} \end{cases} .$$

which means that $Y_{i*} \cap Y_i \neq \phi$.

IV. Regression Method with Kernel

We extend our linear model to a regression model with a kernel-like function, we call modified kernel, on a set of L-R fuzzy number. First we describe a general formula, then give more precise formula as an extension of the polynomial kernel, $k_P(x, y)$, for the case of degree $d = 2$ and the number of explanatory variables $k = 3$ as described in B. of section II.

A. General Formula

We suppose that there exists a function $K(X, Y)$ satisfying only $K(Y, X) = K(X, Y)$ on the product of a set of fuzzy numbers, $\mathcal{X}_F^k \times \mathcal{X}_F^k$ to $\mathcal{X}_F$. Actually, we use a function induced from one of kernels explained in A. of section II if it can be well-defined on fuzzy numbers.

For a given data set of L-R fuzzy numbers, $\{(Y_i, \mathbb{X}_i); i = 1, \ldots, M\}$, where $Y_i = (y_i; p_i, q_i)_F$, $\mathbb{X}_i = (X_{i1}, \ldots X_{ik})_F$ with $X_{ij} = (x_{ij}; w_{ij}, z_{ij})_F$ ($i = 1, \ldots, M, j = 1, \ldots, k$). We just modify the formula (12) by replacing X_j with $K(\mathbb{X}, \mathbb{X}_i)$, and consider the model

$$Y = A_1 K(\mathbb{X}, \mathbb{X}_1) + A_2 K(\mathbb{X}, \mathbb{X}_2) + \cdots + A_M K(\mathbb{X}, \mathbb{X}_M), \tag{19}$$

where $\mathbb{X} = (X_1, \ldots, X_k)$ is vector expression of the explanation fuzzy variable and Y is the objective fuzzy variable. For this formula, we can apply our proposed method for the dual model with h-cut. Since the number of data, M, is usually much greater than the number of explanatory variables, k, the possibility of existence for the lower model increases from the Theorem 4.

On the other hand, when M is very big, there will be too many possible fuzzy number coefficients $\{A_i\}$ for both upper and lower model. Thus, try to find smaller set of representer if possible, and denote their number by N. Then fuzzy coefficients $\mathbb{A}^* = (A_1^*, \ldots, A_N^*)$ and $\mathbb{A}_* = (A_{1*}, \ldots, A_{N*})$ are calculated for upper and lower models from the following formulas of fuzzy numbers respectively,

$$A_1 K(\mathbb{X}_i, \tilde{\mathbb{X}}_1) + A_2 K(\mathbb{X}_i, \tilde{\mathbb{X}}_2) + \cdots + A_N K(\mathbb{X}_i, \tilde{\mathbb{X}}_N), \tag{20}$$

where $i = 1, \ldots, M$, and $\{\tilde{\mathbb{X}}_l; l = 1, \ldots, N\}$ is a representer.

B. Case of Modified Polynomial Kernel

Here we consider a modified kernel induced from polynomial kernel, $k_P(x, y)$, denoted by $K_F(\mathbb{X}, \tilde{\mathbb{X}}) = (\mathbb{X}^t \tilde{\mathbb{X}} + C)^d$. When we could find $N (= {}_{k+d}C_d)$ number of proper value vectors $\tilde{\vec{x}}_l = (\tilde{x}_{l1}, \ldots, \tilde{x}_{lk})$ ($l = 1, \ldots, N$) for the dense subspace of $\mathcal{H}_{k_P}$, put $\tilde{\mathbb{X}}_l = (\tilde{X}_{l1}, \ldots, \tilde{X}_{lk})$ with $\tilde{X}_{li} = (\tilde{x}_{li}; 0, 0)_F$ ($l = 1, \ldots, N$).

Now calculate the h-cut of the equation (20) for $C = (c; 0, 0)_F$ in the way of B. of section III. When putting $\vec{x}_i =$

$(x_{i1},\ldots,x_{ik})$, $\vec{w}_i=(w_{i1},\ldots,w_{ik})$, $\vec{z}_i=(z_{i1},\ldots,z_{ik})$, $i=1,\ldots,M$,we have

$$[\mathbb{X}_i]_h=([X_{i1}]_h,\ldots,[X_{ik}]_h)=[\vec{x}_i-L^{-1}(h)\vec{w}_i,\vec{x}_i+R^{-1}(h)\vec{z}_i],$$

and the h-cut of the modified kernel is as follows,

$$\begin{aligned}[K(\mathbb{X}_i,\tilde{\mathbb{X}}_l)]_h&=\left(\textstyle\sum_j^k[X_{ij}]_h[\tilde{X}_{lj}]_h+[C]_h\right)^d\\&=\left[\left(\textstyle\sum_{j=1}^k(x_{ij}-w_{ij}L^{-1}(h))\tilde{x}_{lj}+c\right)^d,\right.\\&\qquad\left.\left(\textstyle\sum_{j=1}^k(x_{ij}+z_{ij}R^{-1}(h))\tilde{x}_{lj}+c\right)^d\right]\\&=\left[k_P(\vec{x}_i-L^{-1}(h)\vec{w}_i,\vec{\tilde{x}}_l),k_P(\vec{x}_i+R^{-1}(h)\vec{z}_i,\vec{\tilde{x}}_l)\right].\end{aligned}$$

Thus we have

$$\begin{aligned}&\left[\sum_{l=1}^N A_lK(\mathbb{X}_i,\tilde{\mathbb{X}}_l)\right]_h\\&=\left[\sum_{l=1}^N(a_l-r_lL^{-1}(h))k_P(\vec{x}_i-L^{-1}(h)\vec{w}_i,\vec{\tilde{x}}_l),\right.\\&\qquad\left.\sum_{l=1}^N(a_l+s_lR^{-1}(h))k_P(\vec{x}_i+R^{-1}(h)\vec{z}_i,\vec{\tilde{x}}_l)\right],\end{aligned}\tag{21}$$

and minimizing or maximizing objective value is

$$\begin{aligned}J(\mathbb{A})=&\sum_{l=1}^N a_l\left(\tfrac{1}{M}\textstyle\sum_{i=1}^M\left(k_P(\vec{x}_i+R^{-1}(h)\vec{z}_i,\vec{\tilde{x}}_l)\right.\right.\\&\qquad\left.\left.-k_P(\vec{x}_i-L^{-1}(h)\vec{w}_i,\vec{\tilde{x}}_l)\right)\right)\\&+R^{-1}(h)\sum_{l=1}^N s_l\tfrac{1}{M}\textstyle\sum_{i=1}^M k_P(\vec{x}_i+R^{-1}(h)\vec{z}_i,\vec{\tilde{x}}_l)\\&+L^{-1}(h)\sum_{l=1}^N r_l\tfrac{1}{M}\textstyle\sum_{i=1}^M k_P(\vec{x}_i-L^{-1}(h)\vec{w}_i,\vec{\tilde{x}}_l),\end{aligned}\tag{22}$$

where $\vec{\tilde{x}}_l=(\tilde{x}_{l1},\ldots,\tilde{x}_{lk})$ for $l=1,\ldots,N$.

Then our upper model Y^* is constructed with $A_j^*=(a_j^*;r_j^*,s_j^*)_F$ minimizing $J(\mathbb{A}^*)$ under the condition that for all $i=1,\ldots,M$,

$$\begin{cases}y_i-p_iL^{-1}(h)\geq\sum_{l=1}^N(a_l^*-r_l^*L^{-1}(h))\times\\\qquad k_P(\vec{x}_i-L^{-1}(h)\vec{w}_i,\vec{\tilde{x}}_l)\\y_i+q_iR^{-1}(h)\leq\sum_{l=1}^N(a_l^*+s_l^*R^{-1}(h))\times\\\qquad k_P(\vec{x}_i+R^{-1}(h)\vec{z}_i,\vec{\tilde{x}}_l)\\r_j^*,\ s_j^*\geq 0\end{cases}.\tag{23}$$

The lower model Y_* is similarly constructed with $A_{j*}=(a_{j*};r_{j*},s_{j*})_F$ maximizing $J(\mathbb{A}_*)$ under the condition that for all $i=1,\ldots,M$,

$$\begin{cases}y_i-p_iL^{-1}(h)\leq\sum_{l=1}^N(a_{l*}-r_{l*}L^{-1}(h))\times\\\qquad k_P(\vec{x}_i-L^{-1}(h)\vec{w}_i,\vec{\tilde{x}}_l)\\y_i+q_iR^{-1}(h)\geq\sum_{l=1}^N(a_{l*}+s_*lR^{-1}(h))\times\\\qquad k_P(\vec{x}_i+R^{-1}(h)\vec{z}_i,\vec{\tilde{x}}_l)\\r_{j*},\ s_{j*}\geq 0\end{cases}.\tag{24}$$

We also have the same kind of theorem as Theorem 4.

Theorem 5. *When $k_P(\vec{x}_i-L^{-1}(h)\vec{w}_i,\vec{\tilde{x}}_l)>0$ and $k_P(\vec{x}_i+R^{-1}(h)\vec{z}_i,\vec{\tilde{x}}_l)>0$ ($i=1,\ldots,M,\ l=1,\ldots,N$), then*

1) *There always exists an optimal solution in the upper regression model under (23) .*
2) *There exists an optimal solution in the lower regression model under (24) if and only if there exist $a_{1*}^{(0)},\ldots,a_{N*}^{(0)}$ satisfying*

$$\begin{cases}y_i-p_iL^{-1}(h)\leq\sum_{l=1}^N k_P(\vec{x}_i-L^{-1}(h)\vec{w}_i,\vec{\tilde{x}}_l)a_{l*}^{(0)}\\y_i+q_iR^{-1}(h)\geq\sum_{l=1}^N k_P(\vec{x}_i+R^{-1}(h)\vec{z}_i,\vec{\tilde{x}}_l)a_{l*}^{(0)}\end{cases}.\tag{25}$$

C. Illustrative Example

As an illustrative example, we consider a polynomial kernel $k_P(x,y)$ of degree $d=2$ and the number of explanatory variables $k=3$ cases, so the number of basis for the dense subspace $\mathcal{H}_k'$ of $\mathcal{H}_k$ is $N=10$. Only considering triangular type fuzzy numbers, i. e., $L=R$ is the linear function from $(0,1)$ to $(1,0)$ and $L^{-1}(h)=R^{-1}(h)=1-h$, and using the base vectors given in B. of section II, we have

$$\begin{aligned}&\tilde{\mathbb{X}}_l=(\tilde{X}_{l1},\tilde{X}_{l2},\tilde{X}_{l3})\quad(l=1,\ldots,10)\text{ with}\\&\tilde{X}_{11}=(1;0,0)_F,\tilde{X}_{22}=(1;0,0)_F,\tilde{X}_{33}=(1;0,0)_F,\\&\tilde{X}_{41}=(-1;0,0)_F,\tilde{X}_{52}=(-1;0,0)_F,\tilde{X}_{63}=(-1;0,0)_F,\\&\tilde{X}_{71}=(1;0,0)_F,\tilde{X}_{72}=(1;0,0)_F,\\&\tilde{X}_{82}=(1;0,0)_F,\tilde{X}_{83}=(1;0,0)_F,\\&\tilde{X}_{91}=(1;0,0)_F,\tilde{X}_{93}=(1;0,0)_F,\\&\tilde{X}_{lj}=(0;,0,0)_F\quad\text{otherwise.}\end{aligned}$$

Here, we have $M=8$ pairs of fuzzy numbers as an example data set shown in Table I. From these fuzzy numbers, calculate $k_P(\vec{x}_i-L^{-1}(h)\vec{w}_i,\vec{\tilde{x}}_l)$ and $k_P(\vec{x}_i+R^{-1}(h)\vec{z}_i,\vec{\tilde{x}}_l)$ for each pair of (i,l) $(i=1,\ldots,8,\ l=1,\ldots,10)$, then take averages through i for each l. Notice that the calculation is done using $\vec{\tilde{x}}_l$ not $\tilde{X}_{l,i}$.

Next, after setting the constant value for c and the value for h-cut, solve two LP problems, one is for upper model with $\mathbb{A}^*$ and the other is lower model with $\mathbb{A}_*$, satisfying the conditions (23) and (24) respectively.

TABLE I. DATA SET FOR THE ILLUSTRATIVE EXAMPLE

$(y;p,q)_F$	$(x_1;w_1,z_1)_F$	$(x_2;w_2,z_2)_F$	$(x_3;w_3,z_3)_F$
$(3.5;1.5,1.5)$	$(1.0;0.5,0.1)$	$(2.0;0.5,0.5)$	$(3.0;0.5,1.0)$
$(4.5;2.0,2.0)$	$(2.0;0.5,0.1)$	$(2.0;0.5,1.0)$	$(3.5;0.75,1.0)$
$(7.0;2.5,2.5)$	$(3.0;0.1,0.0)$	$(6.5;0.5,1.5)$	$(5.5;1.0,1.25)$
$(9.5;2.0,2.0)$	$(2.0;0.5,0.1)$	$(9.5;1.0,0.5)$	$(10.0;2.0,2.5)$
$(11.0;3.0,3.0)$	$(4.0;0.5,1.0)$	$(9.0;1.0,1.0)$	$(10.5;3.0,2.5)$
$(6.0;2.0,2.0)$	$(2.0;0.0,0.0)$	$(3.0;1.0,2.0)$	$(2.0;0.5,1.0)$
$(8.0;2.5,2.5)$	$(3.0;0.1,0.0)$	$(5.0;1.5,1.5)$	$(5.0;1.5,2.0)$
$(9.0;3.0,3.0)$	$(3.5;0.5,0.0)$	$(4.0;0.5,0.5)$	$(6.0;2.0,1.25)$

By applying the solver function in MS-EXCEL, when setting $c=1$ and $h=0.3$, for the upper model we have

$$\begin{aligned}&A_1^*=(0.218;0,0.038)_F,A_6^*=(0.030;0,0)_F,\\&A_{10}^*=(1.455;0,5.230)_F,A_l^*=(0;0,0)_F\quad\text{(for other }l),\end{aligned}$$

and

$$Y=A_1^*K(\mathbb{X},\tilde{\mathbb{X}}_1)+A_6^*K(\mathbb{X},\tilde{\mathbb{X}}_6)+A_{10}^*K(\mathbb{X},\tilde{\mathbb{X}}_{10}).\tag{26}$$

For the lower model, we have

$$A_{1*} = (0.160; 0, 0.038)_F, A_{2*} = (0.037; 0, 0)_F,$$
$$A_{3*} = (0.002; 0, 0)_F, A_{10*} = (3.301; 0, 0.167)_F,$$
$$A_{l*} = (0; 0, 0)_F \quad \text{(for other } l\text{)},$$

and

$$Y = A_{1*}K(\mathbb{X}, \tilde{\mathbb{X}}_1) + A_{2*}K(\mathbb{X}, \tilde{\mathbb{X}}_2) + A_{3*}K(\mathbb{X}, \tilde{\mathbb{X}}_3) + A_{10*}K(\mathbb{X}, \tilde{\mathbb{X}}_{10}). \quad (27)$$

Table II describes the correspondence of original values and the resulted values by lower model (27) and by upper model (26). The expression of fuzzy numbers here is not the same as used so far in this paper. These values express the left edge, the center point, and the right edge of each triangular shape. We can see three corresponding fuzzy numbers have no inclusion relation, because they are full numbers before operating h-cut procedure. When looking at the support interval of h-cut of each fuzzy set, we have the set relationship $[Y_*]_h \subset [Y]_h \subset [Y^*]_h$. Figure 1 illustrates the relationship among three fuzzy numbers from the second row in Table II.

TABLE II. COMPARISON: Y, Y^*, AND Y_*

$(y-p, y, y+q)$	(y^*-p^*, y^*, y^*+q^*)	(y_*-p_*, y_*, y_*+q_2)
$(2.0, 3.5, 5.0)$	$(2.0, 2.4, 8.1)$	$(3.9, 4.3, 4.8)$
$(2.5, 4.5, 6.5)$	$(2.9, 3.6, 9.5)$	$(4.6, 5.1, 6.0)$
$(4.5, 7.0, 9.5)$	$(5.1, 5.6, 11.8)$	$(7.6, 8.0, 9.8)$
$(7.5, 9.5, 11.5)$	$(4.3, 5.8, 13.1)$	$(7.8, 9.1, 10.2)$
$(8.0, 11.0, 14.0)$	$(7.1, 9.6, 20.2)$	$(9.7, 11.3, 15.5)$
$(4.0, 6.0, 8.0)$	$(3.4, 3.4, 9.1)$	$(5.1, 5.4, 6.6)$
$(5.5, 8.0, 10.5)$	$(5.0, 5.4, 11.9)$	$(6.5, 7.3, 8.8)$
$(6.0, 9.0, 12.0)$	$(5.2, 6.6, 13.0)$	$(6.7, 7.6, 8.7)$

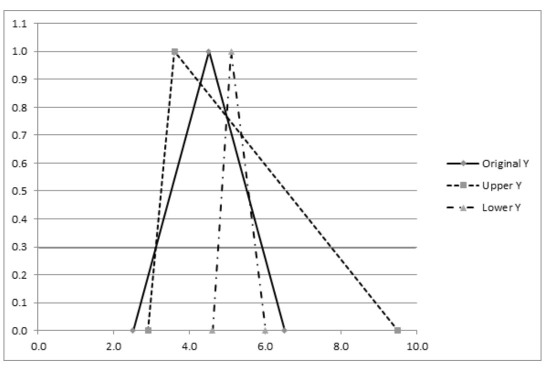

Figure 1. Relationship of Three Fuzzy Numbers

We also tried other type of kernels for these test data, and may have some discussion on the fitness.

V. CONCLUSION

As an extension of our fuzzy dual linear regression model, we proposed to apply kernel method and give a general formula with a modified kernel of polynomial type. Then, we showed how it works using artificial sample data set for illustration of performance in a simple case.

Although we could see that the kernel method can be incorporated with fuzzy regression model, the effectiveness of our method, depending on data set type, is not yet clear. In the example handling small data, when changing the values slightly, we could not have any solution for the lower model. This infeasibility also occurs by increasing the value of h, which may reduce the degree of freedom of resulted fuzzy number of lower model. Though the number of data is less than the number of base set, the merit of choosing base set is that the number N depends only on the degree of kernel and the number of explanatory variables, and does not depend on the size of data set, M.

In order to apply our model to real data, we need to prepare several types of modified kernel model and need to investigate feasibility conditions for the induced LP problem.

REFERENCES

[1] M. A. Aizerman, E. M. Braverman, and L. I. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and Remote Control*, Vol. 25, pp. 821837, 1964.

[2] A. Alim, F. T. Johora, S. Babu, and A. Sultana, "Elementary Operations on L-R Fuzzy Number," *Advanced in Pure Mathematics*, 5, pp. 131-136, 2015.

[3] M. Amagasa, "Formulation of A Sale Price Prediction Model Based on Fuzzy Regression Analysis," *Operations Research Proceedings 2011*, Springer Berlin Heidelberg, pp 567-572, 2012.

[4] M. Amagasa and K. Nagata, "Prediction Model with Interval Data - Toward Practical Applications-," *Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems(IPMU 2016)*, Part II, CCIS611, pp.213-224, 2016.

[5] N. Aronszajn, "Theory of reproducing kernels," *Transaction of the American Mathematical Society*, 68(3), pp.337-404, 1950.

[6] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *Journal of Machine Learning Research*, Vol. 3, pp.1-48, 2002.

[7] B. E. Boser, I. M. Guyon, V. N. Vapnik, "A training algorithm for optimal margin classifiers," *Proceedings of the fifth annual ACM workshop on Computational learning theory(COLT'92)*, pp. 144-152, ACM Press, 1992.

[8] P. D'Urso, "Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data," *Computational Statistics & Data Analysis*, 42, pp. 47-72, 2003.

[9] B. T. Gale, *Managing Customer Value*, FREE PRESS, 1994.

[10] P. Guo, and H. Tanaka, "Dual Models for Possibilistic Regression Analysis," *Computational Statistics & Data Analysis*, Vol.51(1), pp. 252-266, 2006.

[11] Y. Inoue, Y. Uematsu, M. Amagasa, and G. Tomizawa, "The method of setting forecast selling price model by fuzzy regression analysis," *Proceedings of Japan industrial Management Association*, Autumn Congress, pp.194-195, 1991.

[12] T. Melzer, M. Reiter, and H. Bischof, " Nonlinear feature extraction using generalized canonical correction analysis, " *Proceedings of Internal Conference Artificial Neural Networks(ICANN2001)*, pp.353-360, 2001.

[13] S. Mika, G. Rätsh, J. Weston, B. Schölkopf, and K.-R. Müller, "Fisher's discriminant analysis with kernels," *Neural Networks for Signal Processing,*, Vol. 11, pp. 41-48, IEEE, 1999.

[14] R. Rosipal and L. J. Trejo, "Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space," *Journal of Machine Learning Research*, 2, pp. 97-123, 2001.

[15] B. Schölkopf, R. Herbrich, and A. J. Smola, " A Generalized Representer Theorem," *Proceedings of 14th Annual Conference on Computational Learning Theory and 5th European Conference on Computational Learning Theory*, pp. 416-426, Springer-Verlag, 2001.

[16] B. Schölkopf, A. J. Smola, and K.-R. Müller, " Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Computation*, Vol. 10, No. 5, pp.1299-1319, MIT Press, 2006.

[17] J. Shawe-Taylor and N. Cristinanini, "Margin distribution and soft margin, " *Advanced in Large Margin Classifiers*, pp. 349-358, MIT Press, 2000.

[18] H. Tanaka, S. Uejima, and K. Asai, "Linear regression analysis with fuzzy model," IEEE Transactions on Systems, Man and Cybernetics, SMC-12, No.6, pp.903-907, 1982.

[19] H. Tanaka, "Fuzzy analysis by possibility linear models, " *Fuzzy Sets and Systems*, Vol.24, pp.363-375, 1987.