

Comparison of Fuzzy Functions with Fuzzy Rule Base Approaches

I. Burhan Türkşen, Asli Celikyilmaz

Abstract

“Fuzzy Functions” are proposed to be determined separately by two regression estimation models: the least squares estimation (LSE), and Support Vector Machines for Regression (SVR), techniques for the development of fuzzy system models. LSE model tries to estimate the fuzzy function parameters linearly in the original space, whereas SVR algorithm maps the data samples into higher dimensional feature space and estimates a linear fuzzy function in the feature space. The membership values of input vectors are calculated using FCM algorithm or any variation of it. They are then used with scalar input variables by the LSE and SVR techniques to determine “Fuzzy Functions” for each cluster identified by FCM. “Fuzzy Functions” estimated with LSE and SVR methodologies are proposed as alternate representations and reasoning schemas to the fuzzy rule base approaches. We show with three case studies that the new approaches give better results in comparison to well-known fuzzy rule base approaches, i.e., Takagi-Sugeno [28] and Sugeno-Yasukawa [27] in test cases.

Keywords: fuzzy functions, rule bases, membership values, reasoning, least squares, support vector machines for regression

1. Introduction

Fuzzy Functions (FF) approach with least squares estimation (LSE) specified in this paper is proposed by Türkşen [43] for the structure identification of system models and reasoning with them. These fuzzy functions can be determined by any function identification method such as least squares’ estimates, LSE, maximum likelihood estimates, MLE, support vector machine

estimates, SVM, etc. In this paper, we discuss the **Fuzzy Functions with Least Squares’ Estimates, FF-LSE** [43] and the **Fuzzy Functions with Support Vector Machines (FF-SVM)** [3]. In addition, we discuss only the Type 1 Fuzzy Functions.

The proposed Fuzzy Functions using LSE, (FF-LSE) and Fuzzy Functions using SVM (FF-SVM) are formed with a selected set of original scalar input variables plus suitable transformations of membership values of a given input vector belonging to a fuzzy cluster. The membership values are determined by a fuzzy clustering algorithm, e.g., FCM [2] with the analysis of training vectors. The models are validated by test vectors.

In what follows is the description of how the proposed structure identification of system models with FF-LSE and FF-SVM models are unique and structurally different and distinct from the well-known structure identification approaches, which are:

(i) Fuzzy rule bases which are determined either by experts or fuzzy clustering methods, such as FCM [2], in order to obtain the membership descriptions of the input fuzzy sets that form the left hand sides and the output fuzzy sets that form the right hand sides of a fuzzy rule base. This approach was initially proposed by Zadeh [47]-[48] which we denote as Zadeh Fuzzy Rule Base (**Z-FRB**), and originally applied by Mamdani, *et al.*[21]. Among many variations of them are well known Sugeno-Yasukawa [27] and Takagi-Sugeno [28] fuzzy system model.

(ii) Fuzzy Regression models: (a) Models that are proposed by Tanaka [29] and investigated by Tanaka, *et al.* [30]-[32], Celmins [6], Savic and Pedrycz [25] and many other variations. These fuzzy regression models are based on the possibility theory instead of the probability theory. (b) Models that are proposed by Hathaway and Bezdek [19] first determine the fuzzy clusters by a Fuzzy C-regression Model (**FCRM**), which defines how many ordinary regressions are to be constructed, one for each local fuzzy partition. Next each fuzzy partition is used essentially for switching purposes to determine the most appropriate linear/non-linear functions.

We should note that, the proposed “Fuzzy Functions-LSE” and “Fuzzy Functions-SVM” fuzzy system models are alternatives to “Fuzzy Rule Base” approaches as well as “Fuzzy Regression” approaches. In the following, we will first describe the general fuzzy

Corresponding Author: I.Burhan Türkşen is Head Department of Industrial Engineering TOBB-Economy and Technology University, Ankara, Turkey and he is also with Department of Mechanical & Industrial Engineering University of Toronto, Canada

E-mail: turksen@mie.utoronto.ca

Asli Celikyilmaz is with Department of Mechanical & Industrial Engineering University of Toronto, Canada,

Email: asli.celikyilmaz@utoronto.ca

system models and Fuzzy Functions briefly. Then, we are going to specify the details of our proposed approaches FF-LSE and FF-SVM. In the experiments section, we will compare our results with the Sugeno-Yasukawa [27] and Takagi-Sugeno [28] approaches to show that the proposed approaches perform better than general fuzzy rule base approaches.

2. Fuzzy System Models

The basic structure of general fuzzy system models is the Fuzzy Rule Bases (FRB), which attempt to identify the underlying relationship between multi-input and multi-output variables of a system by fuzzy sets. Conceptually, a system with multiple independent outputs can be considered several groups of single output systems separately. Consequently, the general rule structure of multi-input-multi-output (MIMO) fuzzy system can also be presented as a collection of multi-input-single-output (MISO) fuzzy system such that for a single system with an s outputs each multi-consequent rule is broken down into s single-consequent rules. This will increase the number of rules in the system; however the modeling and inference would be less complex for MISO models. Therefore, in this paper, we will only deal with Multi-Input Single Output (MISO) systems. Generally, fuzzy system models represent relationships between the input and output variables, which are expressed as a collection of IF-THEN rules that, utilize linguistic labels, which are represented with fuzzy sets. The general fuzzy rule base structure can be written as follows:

$$R : \text{ALSO}_{i=1}^{c^*} (\text{IF } \textit{antecedent}_i \text{ THEN } \textit{consequent}_i) \quad (1)$$

where c^* is the number of rules in a rule base either given by experts or it is determined by a fuzzy clustering algorithm such as FCM [2], $1 < i < c^*$ and R represents the rule base structure of c^* rules. The calculation of the output from the rule base combines the identification of the inferred output fuzzy set for each rule by an implication operator/connective, "IMP", and aggregation of the deduced output fuzzy sets of each rule with an "ALSO" connective. During the aggregation of the rules, the "ALSO" connective is taken as a t-norm if t-conorm is used in place of "IMP" connective or is taken as a t-conorm if a t-norm is used in place of "IMP" connective [18].

The fuzzy rule base structures determined by alternative (i), which is stated in the introduction section, mainly differ in the representation of the consequents in structure. We will mainly deal with a modified version of these models proposed by (a) Sugeno and Yasukawa [27] which we will call Sugeno-Yasukawa Fuzzy Rule Base structure (**SY-FRB**), and (b) Takagi-Sugeno [28] FRB

structure, which we denote (**TS-FRB**) where the consequents are represented with linear functions of input variables. Next we describe the formalization of SY-FRB and TS-FRB structures.

In general, let nv be the number of selected input variables in the system. Then, the multidimensional antecedent, x , can be defined as $x = (x_1, x_2, \dots, x_{nv})$, where x_j is the j^{th} input variable of the antecedent and the domain of input variables, X , can be defined as $X = X_1 \times X_2 \times \dots \times X_{nv}$, where $X_j \subseteq \mathcal{R}$ is the domain of variable x_j . Similarly, the domain of the output variable, y , will be denoted as $Y \subseteq \mathcal{R}$. Let the i^{th} rule be denoted as R_i , and rule-base with R .

The **SY-FRB** structure can be defined as:

$$R_i : \text{IF } \textit{AND}_{j=1}^{nv} (x_j \in X_j \text{ is } A_{ij}) \text{ THEN } y \in Y \text{ is } B_i, \forall i = 1, \dots, c^* \quad (2)$$

$$R : \text{ALSO}_{i=1}^{c^*} \left(\text{IF } \textit{AND}_{j=1}^{nv} (x_j \in X_j \text{ is } A_{ij}) \text{ THEN } y \in Y \text{ is } B_i \right) \quad (3)$$

where A_{ij} is the linguistic label, i.e., fuzzy subset, associated with j^{th} input variable of the antecedent in the i^{th} rule, R_i , with membership function $\mu_i(x_j) : X_j \rightarrow [0, 1]$ and similarly B_i is the consequent linguistic label, i.e., consequent fuzzy subset, of the i^{th} rule with membership function $\mu_i(y) : Y \rightarrow [0, 1]$, and c^* is the number of rules in the model.

Some of the challenges of the SY-FRB structure are: identification of the membership functions of fuzzy sets on the left and right hand sides of the rules, and identification of the most suitable t-norm and t-conorm combinations that correspond to the linguistic "AND" in FDCF, Fuzzy Disjunctive Canonical Form, and "OR" in FCCF, Fuzzy Conjunctive Canonical Form, for the combination of left hand side fuzzy subsets together with the implication operator, "IMP" in FCCF, Fuzzy Conjunctive Canonical Form, that will carry the left hand side membership degree, i.e., the degree of firing, to the right hand side consequent fuzzy subset.

Another problem arises when one cannot, or one chooses not to, represent the linguistic "AND", in FDCF only, and "OR", in FCCF only, in a one-to-one correspondence with a t-norm and a t-conorm, respectively, as it is shown by Türkşen [33-42]. In such cases the FDCF and FCCF, Fuzzy Disjunctive and Conjunctive Canonical Forms, are both to be used to capture the uncertainty associated with the linguistic "AND", "OR" and "IMP" for the representation of rules and for reasoning with them. Hence, such models fall into Interval-Valued Type 2 fuzzy systems analyses, which we are not dealing with in this paper. Finally, one has to carry out defuzzification computations in all fuzzy rule base models.

The structure above assumes non-interactivity

between input variables [48]. In fact, this is the underlying assumption when the fuzzy subsets for the left and right hand sides are obtained from experts by interview techniques or by projection of membership values into variable space dimensions to determine membership functions after one determines membership values with FCM. In order to eliminate the non-interactivity assumption, Delgado *et. al.* [11], Babuska *et. al.* [1], and Uncu and Türkşen [44] used multi-dimensional Type 1 fuzzy subsets to represent the antecedent part of the rules.

On the other hand, the **TS-FRB** can be expressed as follows:

$$R : \underset{i=1}{c^*} \text{ (IF antecedent}_i \text{ THEN } y_i = a_i x^T + b_i) \quad (4)$$

where, $\text{antecedent}_i \rightarrow x \in X$ is A_i , and $a_i = (a_{i,1}, \dots, a_{i,m})$ is the regression coefficient associated with the i^{th} rule in equation (4) whereas b_i 's are the intercepts associated with the i^{th} rule of equation (4).

Each degree of firing, d_i , associated with the i^{th} rule, is determined directly from the corresponding i^{th} multi-dimensional antecedent fuzzy subset A_i and applied to the consequent fuzzy subset for the SY-FRB or to the classical ordinary regression for the case of TS-FRB.

3. Fuzzy Functions

Although there are many variations of fuzzy functions in the literature [16],[17],[46], a conceptual origin of our proposed FF-LSE may be found in Demirci [12]-[14] where the general properties of "Fuzzy Functions" from the perspective of mathematical theory is discussed.

In particular, he suggests [13] that a possible application of fuzzy functions should be as follows: Let X and Y stand for the universe of discourse of the set of all possible values of the input and output variable sets, respectively. The functional relation between any x and y , which is conceived as an ordinary function $f: X \rightarrow Y$ denotes the functional dependency between x and y . Let the values x -measured and y -measured represent the empirical data and the aim is that the y -measured should coincide with the theoretically expected value $f(x\text{-measured})$ of y . Since, whenever a physical measurement is made, there is always some uncertainty about it, these two values are not usually equal: they can be very close or very similar to each other.

The classical assumption, asserting the functional dependency of y to x as an ordinary function $f: X \rightarrow Y$, does not tell us anything about how the measured value x -measured of x relates to the measured value y -measured of y and how the measured value y -measured of y and the hypothetically claimed value $f(x\text{-measured})$ of y relate to each other. Demirci [13] states that the

main difficulty behind these two problems in the classical approach results from the assumption that each possible value of x is related to a unique possible value of y , and both the indistinguishability of the input values and the indistinguishability of the output values are always omitted mathematically. Instead of the classical assumption that accepts the functional dependency of y to x as an ordinary function $f: X \rightarrow Y$, Demirci [13] proposes a vaguely defined function from X to Y for the description of the functional dependency of y to x to solve these problems.

In other words, taking the M -equivalence relations E on X and F on Y into account, Demirci [13] suggests that a strong fuzzy function $\rho \in L^{X \times Y}$ from X to Y w.r.t. E and F can be taken as the mathematical representation of the functional dependency of y to x , where the M -equivalence relations on X is called an M -equivalent similarity relation [20] such that $M = (L, \leq, *)$ denotes an integral, commutative cqm-lattice (complete quasi monoidal lattice) with $L = [0, 1]$ and $*$ denotes a t-norm. For the proposed ordinary function, $f: X \rightarrow Y$, to represent the functional dependency of y to x , f can be thought as a hypothetical or an ideal description of the functional dependency of y to x . For M -equivalence relations E on X and F on Y , a strong fuzzy function $\rho \in L^{X \times Y}$ from X to Y w.r.t. E and F with the property f in $ORD(\rho)$ (\therefore The set of all ordinary descriptors of ρ is denoted by $ORD(\rho)$) can be also conceived as a realistic and a comprehensive description of the functional dependency of y to x which eliminates the indistinguishability of the input and the output values individually.

For each x in X and y in Y , the element $\rho(x, y) \in L^{X \times Y}$ can be interpreted as **the degree of the truth of the statement**: "y takes the value y for a given value x of x ", where the top element 1 (the bottom element 0) of L denotes the completely true (false) case of this statement. For the sake of simplicity, if we denote the output variable y by $y(x)$ whenever x takes the value x in X , then, for each x in X and y in Y , $\rho(x, y)$ will be nothing but the degree of the truth of " $y(x) = y$ " [13].

It is to be noted that while such theoretical analyses give us a base to start the formation of "Fuzzy Functions", it does not specify how we are to obtain such "Fuzzy Functions". In this regard, our proposed approach, to be discussed next, provides a novel way where such fuzzy functions can be determined in practical engineering applications.

4. Comparison of Proposed Fuzzy Functions Methods with FRB and Fuzzy C-Regression Models (FCRM)

The proposed FF-LSE and FF-SVM's are structurally

different from Z-FRB[47]-[48], SY-FRB [27], TS-FRB [28], and “Fuzzy Regression” models of Tanaka, *et. al.* [32], and its variations which are explained above, and Hathaway and Bezdek [19] model, because the proposed approach introduces membership values and their transformations as new input variables in addition to the original scalar input variables for fuzzy function estimations. But first one executes a fuzzy clustering algorithm such as FCM with original selected input variables after an execution of a feature selection algorithm; and then determines (local) optimum number of fuzzy clusters and hence the associated membership values. Then a fuzzy function to represent each fuzzy cluster separately can be identified. Thus there are as many fuzzy functions as there are fuzzy clusters similar to Hathaway and Bezdek’s [19] Fuzzy C-regression model (FCRM), but they use membership values as the weights to be used in the estimation of the functions using weighted least squares algorithm. FCRM updates the membership values as the similarity measure by using estimation error from these functions.

In this paper, we estimate these fuzzy functions after we learn the membership values of each cluster from FCM algorithm. Therefore it is structurally a new and unique approach for the determination of fuzzy functions instead of fuzzy rule bases. Fuzzy functions represent fuzzy rule bases indirectly. When the relationship between input variables and the output variable of the system can be linearly explained in the original dimension space of the data, it is quite reasonable and faster to estimate the fuzzy functions using least squares estimation. When this relationship is more complex and there needs to be a non-linear transformation of the original input variables, it is better to map the input dataset into a higher dimensional space, e.g., a *hyper-space* where the input dimension is infinite (maximum n). One of the powerful methods to find the fuzzy functions which defines a linear relation between input and output variables in the higher dimension, but a non-linear relationship in the original dimension is the support vector machines which was first proposed by Vapnik [45]. For the regression cases, Support Vector machines for regression algorithm can be applied to find the fuzzy functions. Hence, in the next sections, we are going to specify the details of the Fuzzy Functions estimated using the LSE and Support Vector Machine for Regression, SVR, algorithms.

It is to be noted for the sake of emphasis that the estimated parameters of the inputs, i.e., the coefficients of the inputs in LSE models whether they be membership values and/or their transformations and/or original input variables or the Lagrange multipliers which form the weights for the support vector regression models, are not fuzzy sets in our proposed approach.

Instead membership values and their transformations are augmented into the input set as new and additional variables. In our experience, it is found that this approach is most suitable for those analysts who are familiar with a function estimation technology, e.g., the least squares technology, support vector machines, ridge regression, etc. They only need to develop an understanding of fuzzy clustering algorithms without studying many aspects of fuzzy theory. All they have to understand is the notion of membership values and how they can be obtained from a fuzzy clustering algorithm such as FCM in addition to their usual background knowledge of a function estimation technique, e.g., LSE, or SVR, etc.

Thus we propose a novel approach in order to provide an easy entry into fuzzy system modeling for mathematicians and statisticians who are working in industry and for other novices. For this purpose, we present next our generalization of the LSE algorithm, which includes membership values and their transformations in addition to the original scalar input variables.

5. Proposed Fuzzy Functions Approaches

A. Proposed Fuzzy Functions with LSE (FF-LSE) Method

In ordinary LSE (OLSE) method, the dependent variable, y , is assumed to be a linear function of one or more independent, input, variables, x , plus an error component as follows:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_{nv} x_{nv} + \varepsilon \quad (5)$$

where y is the dependent output, x_j 's are the inputs (explanatory variables), for $j = 1, \dots, nv$, nv is the number of selected inputs and ε is the independent error term which is typically assumed to be normally distributed. The goal of the least squares method is to obtain estimates of the unknown parameters, β_j 's, $j=0,1,\dots, nv$, which indicate how a change in one of the independent variables affects the dependent variable as follows:

$$\beta = (X^T X)^{-1} X^T y \quad (6)$$

where $\beta = (\beta_0, \beta_1, \dots, \beta_{nv})$. The proposed generalization of OLSE as **FF-LSE**, requires that a fuzzy clustering algorithm, such as FCM [2], be available to determine the interactive (joint) membership values of input-output variables in each of the fuzzy clusters that can be identified for a given training data set.

Let (X_k, Y_k) , $k = 1, \dots, nd$, be the set of observations in a training data set, such that

$$X = (x_{j,k} | j = 1, \dots, nv, k = 1, \dots, nd) \quad (7)$$

First, one determines the optimal (m^*, c^*) pair for a particular performance measure, i.e., a cluster validity

index, with an iterative search and an application of FCM algorithm, where m is the level of fuzziness (in our experiments we usually take $m = 1.1, \dots, 2.5$), and c is the number of clusters (in our experiments we usually take $c = 2, \dots, 10$). The well known FCM algorithm can be stated as follows:

$$\begin{aligned} \min J(U, V) &= \sum_{k=1}^{nd} \sum_{i=1}^c (u_{ik})^m (\|x_k - v_i\|_A) \\ \text{s.t.} \quad 0 &\leq u_{ik} \leq 1, \forall i, k \\ \sum_{i=1}^c u_{ik} &= 1, \forall k \\ 0 &\leq \sum_{k=1}^{nd} u_{ik} \leq nd, \forall i \end{aligned}$$

where J is objective function to be minimized, $\|\cdot\|_A$ is a norm that specifies a distance-based similarity between the data vector x_k and a fuzzy cluster center v_i . In particular, $A = I$ is the Euclidian Norm and $A = Cov^{-1}$ is the Mahalonobis Norm, etc., where Cov is the covariance matrix.

The optimal pair, (m^*, c^*) , can be determined with a user defined cluster validity index, et.c, partition entropy or partition coefficient [24]. Another alternative of selecting the optimum pair would be running the overall FF-LSE model for every (m, c) pair specified by the user and determining the optimal pair from the training RMSE values of each model. The following definitions adopt the idea of using a user defined cluster validity index for the determination of the optimal pair. The experiments in this paper follow the second alternative.

Once the optimal pair (m^*, c^*) is determined with the application of FCM algorithm, one next identifies the cluster centers for $m = m^*$ and each cluster $i = 1, \dots, c^*$ as:

$$v_{X|Y, i}^{m^*} = (x_{1,i}, x_{2,i}, \dots, x_{m^*, i}, y_i) \quad (8)$$

From this, we identify the cluster centers of the ‘‘input space’’ again for $m = m^*$ and $c = 1, \dots, c^*$ as:

$$v_{X, i}^{m^*} = (x_{1,i}, x_{2,i}, \dots, x_{m^*, i}) \quad (9)$$

Next, one computes the normalized membership values of each data sample in the training data set with the use of the cluster center values determined in the previous step. There are generally two steps in these calculations:

(a) First we determine the (local) optimum membership values u_{ik} 's and then determine μ_{ik} 's that are above an α -cut in order to eliminate harmonics generated by FCM as:

$$u_{ik} = \left(\sum_{j=1}^c \left(\frac{\|x_k - v_{X, i}\|}{\|x_k - v_{X, j}\|} \right)^{\frac{2}{m-1}} \right)^{-1}, \quad \mu_{ik} = \{u_{ik} \geq \alpha\}, \quad (10)$$

where μ_{ik} denotes the membership value of the k th vector, $k = 1, \dots, nd$, in the i th rule, $i = 1, \dots, c^*$ and x_k denotes the k th vector.

(b) Next, we normalize them as:

$$\gamma_{ik}(x_k) = \frac{\mu_{ik}(x_k)}{\sum_{i'=1}^c \mu_{i'k}(x_k)} \quad (11)$$

where $\gamma_{ik}(x_k)$'s are the normalized membership values of X -domain in the i th cluster, $i = 1, \dots, c^*$, which in turn indicate the membership values that will constitute as a new input variable in our proposed scheme of function identification for the representation of i th cluster. Let $\Gamma_i = (\gamma_{ik} | i = 1, \dots, c^*; k = 1, \dots, nd)$ be the membership values of X data sample in the i th cluster, i.e., i th rule.

Next we determine a new augmented input matrix of X for each of the clusters, which could take on several forms depending on which *transformation* of membership values we want to or need to include in our system structure identification for our intended system analyses. Examples of possible augmented input matrices are:

$$\begin{aligned} X_i' &= [1, \Gamma_i, X], \text{ or} \\ X_i'' &= [1, \Gamma_i^2, X], \text{ or} \\ X_i''' &= [1, \Gamma_i^2, \Gamma_i^m, \exp(\Gamma_i), X], \text{ etc,} \end{aligned} \quad (12)$$

where X_i' , X_i'' , X_i''' are the augmented input matrices to be used in least squares estimation of a new system structure identification and $\Gamma_i = (\gamma_{ik} | i = 1, \dots, c^*; k = 1, \dots, nd)$. The choice depends on whether we want to or need to include just the membership values or some of their transformations as new input variables in order to obtain the best representation of a system behavior. A new augmented input matrix having a single input variable in the original input space when only membership value itself is augmented to the dataset may look like this:

$$X'_i = [1, \Gamma_i, X] = \begin{bmatrix} 1 & \gamma_{i,1} & x_{i,1} \\ \vdots & \vdots & \vdots \\ 1 & \gamma_{i,nd} & x_{i,nd} \end{bmatrix}$$

Up to this point, in the proposed system modeling approach, we have defined how the augmented input matrix for each cluster could be formed using FCM algorithm. Both the proposed FF-LSE and FF-SVM approaches implement these steps. From this point forward, the estimation of fuzzy functions takes place for each cluster, where one can implement any function estimation methodology, e.g., LSE or SVM. Different approaches are followed in the estimation of fuzzy functions using the augmented matrices. Here we continue to specify the FF-LSE models. In the next section the FF-SVM models will be introduced.

Thus the function of a single input single output model, which includes only the membership value as additional input variable, $Y_i = \beta_{i0} + \beta_{i1}\Gamma_i + \beta_{i2}X$, that represents the i^{th} rule corresponding to the i^{th} interactive (joint) cluster in (Y_i, Γ_i, X) space, would be estimated with FF-LSE approach as follows:

$$\beta_i^* = (X_i'^T X_i')^{-1} (X_i'^T Y_i) \quad (13)$$

where $\beta_i^* = (\beta_{i0}^*, \beta_{i1}^*, \beta_{i2}^*)$ and $X_i' = [1, \Gamma_i, X]$, provided the inverse of covariance, $(X_i'^T X_i')^{-1}$, exists.

The estimate of Y_i would be obtained as:

$$Y_i^* = \beta_{i0}^* + \beta_{i1}^* \Gamma_i + \beta_{i2}^* X \quad (14)$$

The single output value is calculated using each output value one from each cluster and weighting them with their corresponding membership values as follows:

$$Y_i^* = \frac{\sum_{i=1}^{c^*} \gamma_i Y_i^*}{\sum_{i=1}^{c^*} \gamma_i} \quad (15)$$

Within the proposed framework, the general form of the shape of a cluster for the case of a single input variable X_j and for the i^{th} cluster can be conceptually captured by a second order (cone) function when one introduces the square of membership values into the augmented input matrix in the space of $[U \times X \times Y]$ which can be illustrated with a prototype shown in Fig. 1.

In a number of real life case studies, we have in fact found out that generally some second order or exponential functions give a good approximation from amongst some 20 alternatives we have experimented.

Before we specify the details of the proposed FF-SVM method, we first review briefly the background of the support vector machines for regression algorithm.

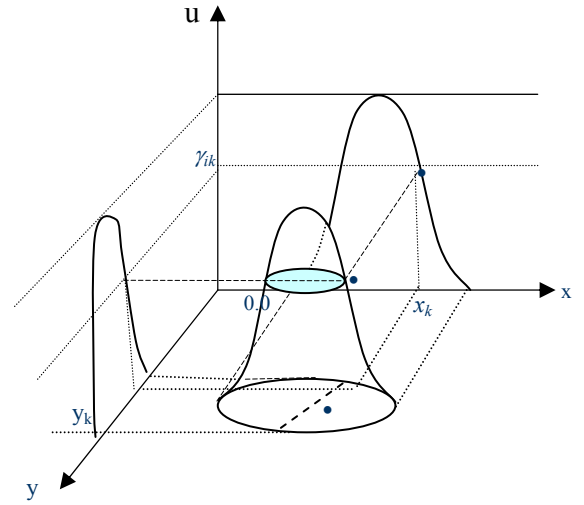


Figure 1. A Fuzzy cluster in $[UxXxY]$ space.

B. Support Vector Machines for Regression

Support Vector Machine (SVM) is a data-mining tool to build a model of a given system. The foundations of SVM have been developed by Vapnik [45]. SVM is a type of optimization technique in which prediction error and model complexity are simultaneously minimized. Let the training samples be denoted as:

$$X = \{(\bar{x}_k, \bar{y}_k) \mid k = 1, \dots, nd\} \quad (16)$$

where X denotes the space of input-output patterns, $\bar{x}_k = (x_{jk} \mid j = 1, \dots, j_{nv}, k = 1, \dots, k_{nd})$ represents each data vector, and y_k is the output value of the k th, data vector in the dataset. Support vector machines are used to solve classification problems as well as regression models where the output variable is scalar. In linear support vector regression (SVR), the aim is to find a pair (\bar{w}, b) , where \bar{w} is the weight vector and b is the bias term in regression equation, such that the value of the point, y_k , can be predicted according to the real-valued function:

$$f(\bar{x}_k) = \hat{y}_k = \langle \bar{w}, \bar{x}_k \rangle + b \quad (17)$$

where $\langle \cdot \rangle$ is the dot product representation. The goal is to find a function, that has at most ε deviation from the actually obtained targets, y_k for all the training data. This concept of ε -insensitive loss function, l_ε , was first introduced by Vapnik [45] as follows:

$$l_\varepsilon = |y_k - f(\bar{x}_k)|_\varepsilon = \max\{0, |y - f(x)| - \varepsilon\} \quad (18)$$

The loss function does not penalize errors below some error, $\varepsilon \geq 0$. Thus the goal of learning is to find a function with a small risk on test samples. This would mean good generalization. This type of SVR is called the ε -insensitive which embodies the Structural Risk

Minimization (SRM) as displayed as follows:

$$R_{exp}[f] \leq R_{emp}[f] + R_{complexity}[f] \quad (19)$$

In SRM of SVM, the aim is to not only minimize the empirical risk from training samples, R_{emp} , but also find a simple function to minimize the complexity of the model, $R_{complexity}$. The more flat the functions are, the less complex they would be, in other words they would get simpler, and therefore they would be closer to the linear functions. The more flat function, the smaller the complexity term of the expected risk in SRM, R_{exp} , gets and the smaller would be the weight vector. In support vector regression the complexity term is expressed as weights assigned to all points in the training sample. In order to ensure that the weight vector is small, Euclidian norm, i.e., $\|w\|^2$ is used. In mathematical terms the objective function of SVM for regression [26] with two conditions can be stated as follows.

$$\begin{aligned} \min_{\bar{w}, b, \xi_k, \xi_k^*} \quad & \zeta(\bar{w}, \xi, \xi^*) = \frac{1}{2} \|\bar{w}\|^2 + \frac{C}{nd} \sum_{k=1}^{nd} (\xi_k + \xi_k^*) \\ \text{subject to} \quad & y_k - \langle \bar{w}, \bar{x}_k \rangle - b \leq \varepsilon + \xi_k \\ & \langle \bar{w}, \bar{x}_k \rangle + b - y_k \leq \varepsilon + \xi_k^* \\ & \xi_k \geq 0 \\ & \xi_k^* \geq 0 \end{aligned} \quad (20)$$

ε represents the ε insensitive value which does not penalize the points whose estimation deviations are lesser/greater than ε , and \bar{w} , b unknown values that represent the weight vector and bias term, respectively. The $C > 0$ determines the tradeoff between the empirical error and the complexity term. The slack variables $\xi_k \geq 0$ and $\xi_k^* \geq 0$ are introduced to the model to soften the optimization problem in order to prevent infeasible solutions. The assumption in (20) is that, it is possible to find such a function that approximates all pairs (\bar{x}_k, y_k) with ε precision. The optimization problem is a convex quadratic program, which can be solved by using the well-known Lagrange Multiplier method. Therefore by introducing Lagrange Multipliers α_i and β_i , one can construct Lagrangian function and the solution to the optimization theorem is given by the saddle point of the Lagrangian function using the Karush-Kuhn-Tucker theorem [23] where the primal model is translated into dual quadratic programming problem as follows:

$$\begin{aligned} \max_{\alpha, \alpha^*} \quad & \frac{1}{2} \sum_{k, k'=1}^{nd} (\alpha_k - \alpha_k^*)(\alpha_{k'} - \alpha_{k'}^*) \langle \bar{x}_k, \bar{x}_{k'} \rangle \\ & - \varepsilon \sum_{k=1}^{nd} (\alpha_k + \alpha_k^*) + \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) y_k \end{aligned}$$

$$\begin{aligned} \text{subject to} \quad & \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) = 0 \\ & \alpha_k, \alpha_k^* \in [0, C] \end{aligned} \quad (21)$$

In model (21), we search for the parameters α_k and α_k^* , which are Lagrange multipliers. The weight vector can now be explained using the Lagrange multipliers as:

$$\bar{w} = \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) \bar{x}_k \quad (22)$$

where, \bar{x}_k is the k^{th} observation and α_k and α_k^* are the Lagrange multipliers for the k^{th} observation, and the estimation function of a vector is given as follows:

$$\begin{aligned} f(\bar{x}_{k'}) &= \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) \langle \bar{x}_{k'}, \bar{x}_k \rangle + b \\ &= \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) \bar{x}_k^T \bar{x}_{k'} + b \end{aligned} \quad (23)$$

where $\bar{x}_{k'}$ is the new vector whose output is to be predicted, T represents the transpose operation on vectors and α_k and α_k^* are the Lagrange multipliers for the k^{th} observation.

In most cases, there is a non-linear relationship between input and output variables and a non-linear support vector regression algorithm is needed. In order to make the support vector model non-linear, the input vectors are mapped into a higher dimensional feature space using a mapping function, $\phi(x)$. However, in most of the cases, explicit mapping results in infeasible solutions and computationally hard. The feasible way to convert a linear SVMs into non-linear SVM is to use kernel mapping which maps the input vectors into a higher dimensional feature space, i.e., $k(x, x') = \langle \phi(x), \phi(x') \rangle$ which changes the SVM algorithm as:

$$\begin{aligned} \max_{\alpha, \alpha^*} \quad & \frac{1}{2} \sum_{k, k'=1}^{nd} (\alpha_k - \alpha_k^*)(\alpha_{k'} - \alpha_{k'}^*) k \langle \bar{x}_k, \bar{x}_{k'} \rangle \\ & - \varepsilon \sum_{k=1}^{nd} (\alpha_k + \alpha_k^*) + \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) y_k \\ \text{subject to} \quad & \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) = 0 \\ & \alpha_k, \alpha_k^* \in [0, C] \end{aligned} \quad (24)$$

Note that one may choose various different kernel functions, e.g., Gaussian Radial Base Kernel, Polynomial Kernel, satisfying the Mercer's condition [26] and the output value of the k' th input vector is calculated using the following function:

$$\hat{y}_k = \hat{f}(\bar{x}_k, \alpha, \alpha^*) = \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) K(\bar{x}_k, \bar{x}_k) + b \quad (25)$$

where the $K(\bar{x}_k, \bar{x}_k)$ represents the kernel mapping of the input vectors. From equation (25), the dependent variable is estimated using the Kernel mapping of the input vectors and their Lagrange multipliers of each vector that are calculated from the optimization algorithm. Note that, (equation (25)) it is not required to calculate the weight vector explicitly, the input vectors whose Lagrange Multipliers are not zero are used in estimating the output and they are called the **support vectors**. In a sense, the complexity of the functions, i.e., the $\frac{1}{2} \|\bar{w}\|^2$ term, represented by support vectors is independent of the dimensionality of the input space X , and only depends on the number of support vectors.

An additional result of the application of Karush-Kuhn-Tucker (KKT) theorem in SVR [23] is:

$$\begin{aligned} (C - \alpha_k) \xi_k &= 0 \\ (C - \alpha_k^*) \xi_k^* &= 0 \end{aligned} \quad (26)$$

One of the several conclusions [23] one might make from (26) is that $\alpha_k \alpha_k^* = 0$, i.e., there can never be a set of dual variables for an observation k which are both simultaneously non-zero as this would require non-zero slacks in both directions. Since $C > 0$, then $\xi_k \xi_k^* = 0$ must also be true. In the same sense, there can never be two slack variables $\xi_k, \xi_k^* > 0$ which are both non-zero and equal.

C. Proposed Fuzzy Functions with SVM (FF-SVM) Method

As one can build ordinary least squares for the estimation of the fuzzy functions when the relationship between input variables and the output variable can be linearly defined in the original input space, one may also build support vector regression models to estimate the parameters of the non-linear fuzzy functions using support vector regression methods [4],[5]. The augmented input matrix is determined from FCM algorithm one for each cluster same as FF-LSE model. One may choose any membership transformation depending on the input dataset. Then one can apply support vector regression algorithm instead of LSE models to each augmented matrix, which are comprised of the original selected input variables and the membership values and/or their transformations. Support vector machines optimization algorithm is applied to each augmented matrix of each cluster (rule) $i, i=1, \dots, c^*$, to optimize their Lagrange multipliers, α_{ik} and α_{ik}^* , and

find the candidate support vectors, $k=1, \dots, nd$. Hence, using FF-SVM, one finds Lagrange multipliers of each k th train data sample one for each cluster, i . Then the output value of k th data sample in i th cluster is estimated using the equation (27) as follows:

$$\hat{y}_{ik}^* = \hat{f}(\bar{x}_{ik}, \alpha_i, \alpha_i^*) = \sum_{k=1}^{nd} (\alpha_{ik} - \alpha_{ik}^*) K(\bar{x}_{ik}, \bar{x}_{ik}) + b_i \quad (27)$$

where the \hat{y}_{ik}^* is the estimated output of the k 'th vector in i th cluster which is estimated using the support vector regression function with the Lagrange multipliers of the i th cluster. The augmented kernel matrix denotes the kernel mapping of the augmented input matrix (as described in FF-LSE approach) where the membership values and their transformations are used as additional input variables. After the optimization algorithm finds the optimum Lagrange multipliers, one can estimate the output value of each data point in each cluster using equation (27).

The inference structure of FF-SVM is adapted from the fuzzy functions with least squares where one can estimate a single output of a data point (see equation (15)) by taking the membership value weighted averages of its output values calculated for each cluster using equation (27).

6. Case Study Applications

In order to test the proposed model performances as opposed to fuzzy rule base systems three input-output datasets are considered in this investigation. These are:

- (i) Daily price of a stock in stock market.
 - (ii) Customer Income Prediction model for a major bank.
 - (iii) The amount of chemicals for a desulphurization process for a steel processing company.
- The specifications of each datasets are displayed in the following parts:

Daily Stock Price Dataset:

Daily stock price dataset comprises of the daily trend data of stock prices. This dataset was introduced by Sugeno and Yasukawa [27]. The same dataset has been used in various other studies one of which compares six different fuzzy reasoning methods using this dataset [23].

Out of 100 observations, 50 of them are used for the training purposes and the other 50 was hold-out for testing purposes. The dataset is given in [23]. There were originally 11 input variables and single output variable in the dataset (see *Appendix Table. I.* for variable definitions). Preliminary input selection was applied using Random Forests (RF) [3] method, which estimates variable importance. Based on the results of RF method,

only 4 of input variables *i.e.*, x_2 , x_4 , x_8 , x_{10} , were found to have importance on the output variable. The rest of the variables had insignificant effect on the output.

Income Prediction Dataset:

The purpose of the Income Prediction Model was to predict the income of the future customers based on the current customer information and 1996 year census data. There were more than 200 variables and hundred of thousands of customers. According to business needs, the dataset was partitioned into 9 different parts based on age, number of different types of investment accounts hold by the customer and different regions of residency. In this study, only one partition was investigated.

We have only selected 10% of the *i.i.d.* data samples to do our research on using only single partition explained above. The data was cleaned from the outliers using the expert's knowledge and 900 training and 900 testing observations are selected randomly. The dataset was comprised of 11 input variables (see *Appendix Table II.*). Based on the correlation analysis, 3 input variables were discarded from the dataset resulting in 8 input variables. There were no census variables among the selected input variables.

Desulphurization Dataset:

A torpedo car desulphurization facility removes sulfur from hot metal leaving the blast furnaces before it is sent to the next process. Generally, desulphurization is carried out by injecting two different powered reagents directly into the hot metal via a lance. The reagents react with the sulfur in the hot metal and residue, which is rich in sulfur, is separated from the iron.

The aim of the data-mining project was to determine the right amounts of the reagents to be added into the hot metal. These reagents are expensive materials and precise estimation is required. There are vast quantities of data available on the desulphurization process, which has various characteristics. The original input and output variables are shown in *Appendix Table III.* There are 750 training and 900 verification samples used in the experiment. Based in the variable selection using random forest regression method [3], only 5 variables are found to be important.

A. Experimental Design

Using the three different input-output datasets, we build four different fuzzy system model structures, FF-LSE, FF-SVM, SY-FRB and TS-FRB. In order to keep the consistency between each model structure, the same training and testing datasets are used for the four fuzzy system models with the same input variables. The categorical variables are transformed into probabilities using logistic regression and are used as additional

inputs only in Income Prediction and Desulphurization Datasets in all of the four models.

B. Sugeno-Yasukawa Models and Takagi-Sugeno Models

The proposed FSM models are compared to two well known Fuzzy Rule Base Models: (i) Sugeno and Yasukawa's [27] fuzzy logic based approach using Partition Type Fuzzy Model, **SY-FRB**, (ii) Takagi and Sugeno's [28] fuzzy system modeling approach, **TS-FRB**. In Sugeno and Yasukawa's FSM approach, they use linguistic measures for both the consequent and the antecedent part of the fuzzy rules and the system learns all inference parameters from the data without the expert intervention. The variable selection method defined in their paper is not applied to these 3 datasets in order to compare the models on the same basis.

In Takagi and Sugeno's FRB (TS-FRB) structure [28], they assume that the antecedent membership functions are to be characterized with triangular membership functions. In their approach, each input variable space is assumed to be partitioned into two clusters and logical connective *AND* is taken as *MIN*. Then, the structure identification problem is just to identify the regression equation coefficients for each rule and the antecedent parameters for each input variable in each rule. Researchers proposed several structure identification methods to identify the membership functions from the data, e.g., Delgado *et.al.*, Babuska and Verbruggen [1], etc. In this paper we have used Babuska *et. al.*'s [1] modified Takagi-Sugeno study where the membership functions of the antecedents are identified using fuzzy c-means clustering and projected onto each input vector. The degree of fulfillment of each rule is then calculated using a t-norm operator, *i.e.* product. Only one aggregate input membership function is identified for each rule. The inference parameters are same as the traditional Takagi-Sugeno inference method [28] where the relationship with the consequents and the antecedents are assumed to come from a linear function.

C. Fuzzy Functions Models

In this paper, the modeling performance from the Fuzzy Functions with LSE and SVM models, FF-LSE and FF-SVM, are compare to the Fuzzy Rule Base structures. Instead of using cluster validity indices to select the optimum model parameters, we measured the optimum model based on the best model performance using RMSE by applying a grid search for each parameter. Note that fuzzy functions with LSE system models have 2 parameters, which are the FCM parameters, *i.e.*, degree of fuzziness and cluster size. Fuzzy Functions with SVR models have 4 parameters: FCM parameters and the SVR parameters which are

C-regularization and ε -insensitive value (epsilon).

In both of these fuzzy function models, membership values and their exponential transformations are used as additional input variables. We applied the LIBSVM program [9] within our fuzzy function codes for support vector optimization in estimating the fuzzy functions. We chose Gaussian RBF as the kernel function, $K(x,x')=exp(-\gamma ||x-x'||^2)$ in all the experiments and the default values of the kernel parameters in LIBSVM [8] are used. As mentioned in section 4.3. the SVR regression has two parameters that is set by the user: ε -insensitive zone (epsilon or ε) and the regularization parameter, C . The parameters of SVM regression, C and ε , are generated by the grid search, i.e., $C=\{2^{-3}, 2^{-1}, \dots, 2^3, 2^5\}$, $\varepsilon=\{0.1, \dots, 0.5\}$, as well as the FCM parameters, i.e., $c=3, 5, \dots, 10$, $m=1.1, \dots, 2.5$ in all the experiments. Hence, for the fuzzy functions with LSE models 2 parameters are specified for each model, and for the fuzzy functions using SVM models, 4 parameters, m , c , C , and γ are determined using a grid search where the model performance of each model is determined using Root Mean Square Error of the models as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \tag{28}$$

y_i , and \hat{y}_i are the actual and estimated output values of a single observation, N is the total number of observations in the dataset.

D. Model Results

The 4 fuzzy system models are applied on the daily stock price of a stock market, income prediction, and reagents estimation in desulphurization process datasets. The results are displayed in Table 1, 2 and 3.

Fuzzy function models, when estimated with either SVM or LSE algorithms, show better generalization than the fuzzy rule base models. The fuzzy function models can increase the model performance by up to 35% depending on the dataset.

Table 1. Daily Stock Price of a Stock in Stock Market

	FF-SVM	FF-LSE	TS-FRB	SY-FRB
RMSE(train)	2.43	3.82	2.76	7.16
RMSE(test)	3.64	5.61	5.68	9.93

Table 2. Income Prediction Dataset*

	FF-SVM	FF-LSE	TS-FRB	SY-FRB
RMSE(train)	0.40	0.52	0.49	0.58
RMSE(test)	0.64	0.64	0.80	0.70

* The RMSE values are calculated from standardized output values.

Table 3. Reagent estimation for Desulphurization Process

	FF-SVM	FF-LSE	TS-FRB	SY-FRB
<i>Reagent1</i>				
RMSE(train)	30	40	35	69.5
RMSE(test)	42	45	45	72
<i>Reagent2</i>				
RMSE(train)	4.80	6.49	5.62	10.01
RMSE(test)	6.59	7.19	7.19	10.80

Table 4. displays the optimum parameters of the models from each experiment whose results are displayed in Table 1-3. In Table 4. *C-reg* indicates the regularization parameter, ε is the ε -insensitive region (epsilon), m refers to the degree of fuzziness (weighting exponent) of the fuzzy c -mean clustering algorithm, c indicates the number of cluster, and #sv refers to the average number of support vectors from each support vector regression model build for each cluster of the FF-SVM models. In order to show the dispersion of the number of support vectors among each cluster we also included the standard deviation (σ_{sv}) of the support vectors of the optimum models.

Table 4. Optimum Model Parameters of three datasets.

Dataset	Model Type	Optimum model Parameters
Daily Stock Price	FF-SVM	<i>C-reg</i> =32, ε =0.2, c =8, m =1.9, #sv=28 (σ_{sv} =1.5)
	FF-LSE	c =8, m =1.6
Income Prediction	FF-SVM	<i>C-reg</i> =64, ε =0.2, c =7, m =1.2, #sv=264 (σ_{sv} =5.7)
	FF-LSE	c =8, m =1.6
Desulphurization	FF-SVM	<i>Reagent 1:</i> <i>C-reg</i> =64, ε =0.2, c =7, m =1.4, #sv=121 (σ_{sv} =4.8)
		<i>Reagent 2:</i> <i>C-reg</i> =64, ε =0.2, c =7, m =1.5, #sv=101 (σ_{sv} =7.7)
	FF-LSE	<i>Reagent 1:</i> c =5, m =1.4
		<i>Reagent 2:</i> c =6, m =1.5

The grid search algorithms applied in this paper try to find the best RMSE value from training data in each experiment and assign these parameters as the optimum model parameters. The algorithm searches for the minimum regression error. Then, verification dataset output is inferred using the optimum parameters. The issue with these grid search algorithms is that, sometimes, the models get stuck in the local minimum which is smaller than the global minimum and this might cause generalization problems. An example to this concept is shown in income prediction dataset (Table 2). The model parameters best fit to the training data when FF-SVM is used but this causes generalization problems. It should also be reminded that, when there is a linear

relationship between the inputs and the output, then LSE model performances will be as good as the other model performances. On the other hand, FF-LSE models, in three of the datasets, show more reliable results than the SVM models. One should run both models and determine the optimum model parameters after observing the results from both models.

7. Conclusions

Two well-known fuzzy rule bases models are compared to two novel fuzzy function approaches. The proposed fuzzy functions are distinct and uniquely different in structure identification and reasoning from: 1) the rule bases originally proposed by Zadeh and initially implemented by Mamdani and their important variations proposed by Sugeno-Yasukawa [27] and Tagaki-Sugeno models [28], and 2) fuzzy regression models originally proposed by Tanaka et al [32] and Hathaway and Bezdek [17].

The proposed fuzzy functions and the rule base approaches are structurally different from each other. In fuzzy functions, membership values and their transformations obtained from a fuzzy clustering algorithm enter into an augmented input matrix together with non-scalar original input variables for function identification exercises in addition to original scalar input variables with Least Squares estimation technique or Support Vector Regression. Whereas in fuzzy rule bases, membership functions of fuzzy sets associated with the original input variables as well as the output variable must be estimated either by experts or by a fuzzy clustering technique for representation. Next for approximate reasoning, system analyst needs to identify combination operators known as t-norms and t-conorms and the associated implication operator as well as needs to implement fuzzification, Generalized Modus Ponens and defuzzification. In fuzzy functions approach, the system analysis only needs to select one estimation function to find the local parameters of the model and the suitable membership transformations to apply to the system to identify the local relationships of the input and output dataset.

To demonstrate the performances of the proposed fuzzy functions, we have investigated one benchmark data set that was used in Sugeno-Yasukawa's models and two data sets that came from two real-life case studies. In these applications, we have compared the results of proposed fuzzy function approach to two well-known fuzzy rule base structures, i.e., Sugeno-Yasukawa FRB [27] and Takagi-Sugeno FRB's [28]. The results have shown that Fuzzy Functions models can predict the model output better than the fuzzy rule bases in most of the datasets.

8. Appendix

Table I. Daily Stock Price Estimation Variables.

Var. Name	Variable Description
$x1$	Past change of moving average (1) over a middle period.
$x2$	Present change of moving average (1) over a middle period
$x3$	Past separation ratio (1) with respect to moving average over a middle period
$x4$	Present separation ratio (1) with respect to moving average over a middle period
$x5$	Present change of moving (2) over a short period
$x6$	Past change of price (1), for instance, change on one day before
$x7$	Present change of price (1)
$x8$	Past separation ratio (2) with respect to moving average, over a short period
$x9$	Past change of moving average (3) over a long period
$x10$	Present separation ratio with respect to moving average over a short period
y	Prediction of stock price

Table II. Income Prediction Dataset Variables.

Symbol	Quantity
<i>WITHDRAWL_AM</i>	The 6-month average amount of withdrawal transactions completed by the customer
<i>SYSTEM_AM</i>	The 6-month average amount of transactions completed through the system by the customer
<i>MNYOUT_ACTIVE</i>	The total month-end market value of all active customer account holdings which are money-out (LOC, Loan, Mortgage, Visa)
<i>TOTBAL_AM</i>	The pro-rated balance held in active Mortgage accounts
<i>MTG_ACTIVE_PBAL_AM</i>	The 6-month average amount of the transactions completed through the branches by the customer
<i>BRANCH_AM</i>	Customer's age in years
<i>AGE_YR</i>	The 6-month average of the number of withdrawal transactions completed by the customer
<i>WITHDRAWL_CT</i>	The 6-month average of the number of transactions completed through the system by the customer
<i>SYSTEM_CT</i>	The number of all active accounts in account family DEM (Demand Deposit) and account sub-family CHQ (Personal Demand Chequing)
<i>DEM_CHQ_ACTIVE_CT</i>	The 6-month average of the number of transactions completed through the branches by the customer
<i>BRANCH_CT</i>	The logistic transformation of the categorical variables.
<i>Probability_LR</i>	The Income of the customer
<i>Income</i>	

Table III. Desulphurization Dataset Variables.

Symbol	Quantity
<i>Start-Sulfur</i>	Starting level of sulfur before desulphurization
<i>KGS</i>	Weight of the batch that consists of iron (tons)
<i>TEMP</i>	Temperature of the hot metal as it leaves from the Blast Furnace.
<i>FB</i>	Measure of fullness of the furnace
<i>Aim-Sulfur</i>	The amount of sulfur that is targeted to remain after desulphurization
<i>Compound 1 to 5</i>	The chemicals measured in the hot metal as they arrive the desulphurization process
<i>Probability_</i>	The logistic transformation of the categorical

LR variables
 Reagent1 Two output variables indicating the reagent amounts
 Reagent2 to be added into the hot metal fro desulphurization
 process.

9. References

- [1] R.Babuska, and H.B. Verbruggen, "Constructing Fuzzy Models by Product Space Clustering," in *H. Hellendoorn and D. Driankov(eds.), Fuzzy Model Identification: Selected Approaches*, pp.53-90, Springer, Berlin, Germany, 1997.
- [2] J.C. Bezdek, "Fuzzy Mathematics in Pattern Classification," *Ph.D. Thesis, applied Mathematics Centre, Cornell University, Ithaca*, 1973.
- [3] Breiman, Leo, "Random Forests", *Machine Learning*, vol. 45, no. 1, pp. 5-32, October 2001.
- [4] A. Celikyilmaz, "Fuzzy Functions with Support Vector Machines.", *M.A.Sc. Thesis*, University of Toronto, 2005.
- [5] A. Celikyilmaz, I.B. Turksen, "Fuzzy Functions with Support Vector Machines.", *Information Sciences, accepted*.
- [6] A.Celmins, "Least Squares Model Fitting to Fuzzy Vector Data," *Fuzzy Sets and Systems*, vol. 22, pp. 245-269, 1987.
- [7] A. Celmins, "Multidimensional Least Squares Model Fitting of Fuzzy Models," *Math. Modeling*, vol. 9, pp. 669-690, 1987.
- [8] Y.-H.O. Chang, B.M. Ayyub, "Reliability Analysis in Fuzzy Regression," *Proc., Annual Conf. of NAFIPS'93*, Allentown, PA., USA, pp.93-97, 1993.
- [9] C. Chang., C. Lin, (2001) , "LIBSVM: a library for support vector machines", Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [10] P.-T. Chang, E.S. Lee, "Fuzzy Linear Regression with Spreads Unrestricted in Sign," *Compt. Math. Appl.*, vol. 28, pp. 61-71, 1994.
- [11] M.R.Delgado, A.F. Gomez-Skennata and F.Martin, "Rapid Prototyping of Fuzzy Models," in *H. Hellendoorn and D. Driankov(eds.), Fuzzy Model Identification: Selected Approaches*, pp. 53-90, Springer, Berlin, Germany, 1997.
- [12] M. Demirci, "Fuzzy Functions and their Fundamental Properties," *Fuzzy Sets and Systems*, vol. 106, pp.239-246, 1999.
- [13] M. Demirci, "Foundations of Fuzzy Functions and Vague Algebra Based on Many-Valued Equivalence Relations, Part I: Fuzzy Functions and their Applications," *I.J. General Systems*, vol. 32, pp. 123-155, 2003.
- [14] M. Demirci, and J. Recasens, " Fuzzy Groups, Fuzzy Functions and Fuzzy Equivalence Relations," *Fuzzy Sets and Systems*, vol. 144, pp. 441-458, 2004.
- [15] P. Diamond, "Fuzzy Least Squares," *Inform. Sci.* vol. 46, pp. 141-157, 1998.
- [16] Dubois, H. Prade, *Fuzzy Sets and Systems: Theory and Applications.*, Academic Press, New York, 1980
- [17] D. Dubois, H. Prade, "Towards Fuzzy Differential Calculus: Part1: Integration of Fuzzy Mapping", *Fuzzy Sets and Systems*, vol. 8, pp. 1-17, 1982.
- [18] Emami, M.R., Turksen, I.B., Goldernberg, A.A., "Development of a Systematic Methodology of fuzzy logic modeling", *IEEE Transactions on Fuzzy Systems*, vol. 6, Issue 3, pp. 346-361, 1998.
- [19] R.J. Hathaway and J.C. Bezdek. "Switching regression models and fuzzy clustering," *In IEEE Transactions on fuzzy systems*, vol.1, no.3, pp.195--203, 1993.
- [20] U. Höhle, "Quotients with Respect to Similarity Relations", *Fuzzy Sets and Systems*, vol. 27, pp. 31-44, 1998.
- [21] E.H. Mamdani, S. Assilian, "An Experiment in Linguistic Syntesis with a Fuzzy Logic Controller," in *E.H. Mamdani, B.R. Gains, "Fuzzy Reasoning and Its Applications"*, Academic Press, 1981.
- [22] M. Mizimoto, "Method of Fuzzy Inference Suitable for Fuzzy Control," *J.of Soc.Instrument Control Engineering*, vol.58, pp. 959-963, 1989.
- [23] Nakanishi H., Türkşen I.B., Sugeno M., "A Review and Comparison of Six Reasoning Methods", *Fuzzy Sets and Systems*, vol. 57, pp. 257-295, 1993.
- [24] Pal, N.R., Bezdek, J.C., "On Cluster Validity for the Fuzzy C-Means Model", *IEEE Transactions on Fuzzy Systems*, vol. 3, no.3, 1995.
- [25] D. Savic and W. Pedryzc, "Evolution of Fuzzy Linear Regression Models," *Fuzzy sets and systems*, vol. 39, pp. 51-63, 1991.
- [26] A.J. Smola,, B. Scholkopf, "A Tutorial on Support Vector Regression", *NeuroColT2 Technical Report Series*, NC2-Tr-1998-030, 1998.
- [27] Sugeno M., Yasukawa T., "A Fuzzy Logic Based Approach to Qualitative Modelling", *IEEE Transactions on Fuzzy Systems*, vol. 1, pp. 7-31, 1993.
- [28] T.Takagi and M.Sugeno, "Fuzzy Identification of Systems and Its Applications to Modeling and Control," *IEEE Transactions on Systems, Man and Cybernetics*, vol.SMC-15, no.1, pp.116-132, 1985.
- [29] H. Tanaka, "Fuzzy Data Analysis by Possibilistic Linear Models," *Fuzzy Sets and Systems*, vol. 24, pp. 363-375, 1991.

- [30] H. Tanaka, H. Ishibuchi, "Identification of Possibilistic Linear Systems by Quadratic Membership Functions of Fuzzy Parameters," *Fuzzy Sets and Systems*, vol. 41, pp. 145-160, 1991.
- [31] H. Tanaka, H. Ishibuchi, S. Yoshikawa, "Exponential Possibility Regression Analysis," *Fuzzy Sets and Systems*, vol. 69, pp. 305- 318, 1995.
- [32] H. Tanaka, S. Vegima, K. Asai, "Linear Regression Analysis with Fuzzy Model," *IEEE Trans. On Systems, Man, Cybernetics*, SMC-2, pp. 903-907, 1982.
- [33] I.B. Türkşen, "Interval Valued Fuzzy Sets Based on Normal Forms," *Fuzzy Sets and Systems*, vol. 20, pp.191- 210, 1986.
- [34] I.B. Türkşen, "Four Methods of Approximate Reasoning with Interval-valued Fuzzy Sets", *Int. J. of Approximate Reasoning*, vol. 3, pp.121-142, 1989
- [35] I.B. Türkşen, "Interval-valued fuzzy sets and compensatory AND," *Fuzzy Sets and Systems*, vol. 51, pp. 295- 307, 1992
- [36] I.B. Türkşen, "Interval valued fuzzy sets and fuzzy connectives," *Interval Computations*, pp. 125-142, 1993.
- [37] I.B. Türkşen, "Interval valued fuzzy sets and fuzzy measures", *Proc. of First Int. Conf. Of NAFIPS*, pp. 317- 321, 1994.
- [38] I.B. Türkşen, "Fuzzy Normal Forms", *Fuzzy Sets and Systems*, vol. 69, pp. 319-346, 1995.
- [39] I.B. Türkşen, "Type I and "Interval-Valued" Type II Fuzzy Sets and Logics", *In P. Wang (Ed.), Advances in Fuzzy Set Theory and Technology*, 1995.
- [40] I.B. Türkşen, "Non-specificity and interval-valued fuzzy sets," *Fuzzy Sets and Systems*, vol. 80 , pp. 87-100, 1996.
- [41] I.B. Türkşen, "Type I and Type II fuzzy system modeling," *Fuzzy Sets and Systems*, vol. 106, pp. 11-34, 1999.
- [42] I.B. Türkşen, "Type 2 representation and reasoning for CWW," *Fuzzy Sets and Systems*, vol. 127, pp. 17-36, 2006.
- [43] I.B. Türkşen, "Fuzzy Functions with LSE," *Applied Soft Computing*, (forthcoming)
- [44] Ö. Uncu and I.B. Türkşen, "A Novel Fuzzy System Modeling Approach: Multidimensional Structure Identification and Inference", *Proc. of Tenth IEEE International Conference on Fuzzy Systems, Melbourne, Australia*, December, pp. 557-562, 2001.
- [45] N.V.Vapnik, *Statistical Learning Theory*, John Wiley and Sons, New York, 1998.
- [46] Z. Yue, Q. Zhong, W. Guangyuan, "Solving Process for a system of first-order differential equations", *Fuzzy Sets and Systems* vol. 95, pp. 333-347, 1998.
- [47] L.A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338-353, 1965.
- [48] L.A. Zadeh, "The Concept of a Linguistic Variable and its Application to Approximate Reasoning," *Information Sciences*, vol.8, pp. 199-249, 1975.



I. Burhan Türkşen received the B.S. and M.S. degrees in Industrial Engineering and the Ph.D. degree in Systems Management and Operations Research all from the University of Pittsburgh, PA. He joined the Faculty of Applied Science and Engineering at the University of Toronto and became Full Professor in 1983. In 1984-1985

academic years, he was a Visiting Professor at the Middle East Technical University and Osaka Prefecture University. Since 1987, he has been Director of the Knowledge / Intelligence Systems Laboratory. During the 1991-1992 academic years, he was a Visiting Research Professor at LIFE, Laboratory for International Fuzzy Engineering, and the Chair of Fuzzy Theory at Tokyo Institute of Technology. During 1996 academic year, he was Visiting Research Professor at the University of South Florida, USA, and Bilkent University, Ankara, Turkey. Since December 2005, he is appointed as the Head of Department of Industrial Engineering at TOBB Economics and Technology University in Ankara, Turkey. He is a Fellow of IFSA, IEEE and WIF, and a member of IIE, CSIE, CORS, IFSA, NAFIPS, APEO, APET, TORS, ACM, etc. He received the outstanding paper award from NAFIPS in 1986, "L.A. Zadeh Best Paper Award" from Fuzzy Theory and Technology in 1995, "Science Award" from Middle East Technical University, and an "Honorary Doctorate" from Sakarya University. He is a Foreign Member, Academy of Modern Sciences. He has published near 300 papers in scientific journals and conference proceedings. His book entitled "An Ontological and Epistemological Perspective of Fuzzy theory" was published by Elsevier, The Netherlands, in January, 2006. (ISBN-10: 0-444-51891-6, ISBN-13: 978-0-444-51891-0)



Asli Celikyilmaz is Ph.D. candidate in Department of Mechanical and Industrial Engineering at the University of Toronto. She received her B.S. degree in Industrial Engineering at Istanbul Technical University and received the M.S. degree in Fuzzy System Modeling in Industrial Engineering at the University of Toronto.

Her current research interests include fuzzy functions, fuzzy decision making, support vector machines, ontology engineering and fuzzy data mining of the web.