Hacettepe Journal of Mathematics and Statistics
Volume 41 (3) (2012), 347 – 359

A REVIEW OF DEVELOPMENTS FROM FUZZY RULE BASES TO FUZZY FUNCTIONS

I. Burhan Turksen*†

Received 03 : 01 : 2012 : Accepted 28 : 08 : 2012

Abstract

We first touch upon a Philosophical Grounding of fuzzy theory expressed by Pierce and Zadeh. Then we review briefly basic and well known fuzzy rule base models and their variations as well as our fuzzy functions with LSE and their enhanced version. We propose a potential future investigation for the basic structure of fuzzy function models generated with an additive effect of membership values and suggest future research for a multiplicative affect of membership values.

Keywords: System structure identification, Fuzzy rule bases, Fuzzy functions with additive components, Future research for fuzzy functions with multiplicative components.

2000 AMS Classification: 03 B 52, 03 E 72, 94 D 05, 46 S 40.

1. Philosophical grounding

It is important to understand an in-depth association of the essential concepts that were treated by Charles S. Peirce and Lotfi A. Zadeh.

Peirce's thesis may be represented as "indeterminacy and determinacy" of "symbols". This view can now be interpreted and expressed with the degree assignment to information granules

Thus in general a symbol S is indeterminate iff $(\exists P)\ \sim (S \text{ is } P \text{ or } S \text{ is } \sim P)$.

Hence, Locke's famous idea of the triangle is stated as:

"It is not the case that a triangle in general is scalene or that is it not scalene".

*I. Burhan Turksen, Ph.D., P.Eng. Fellow: IFSA, IEEE, WIF, Department of Industrial Engineering, TOBB-Economics and Technology University, Soğutözü Cad. No: 43, Soğutözü 06560, Ankara, Turkey. E-mail: `bturksen@etu.edu.tr`

†Director-Knowledge/Intelligence Systems Laboratory, Department of Mechanical & Industrial Engineering, University of Toronto, Toronto, Ontario, M5S 3G8, Canada.
E-mail: `turksen@mie.utoronto.ca`

Peirce's discussion pre-supposes that every symbol is capable of determining an interpretant symbol and that symbols are at least potentially general.

$$(S)\ (\exists P)\ (S \text{ is } P \text{ OR } S \text{ is } \sim P) \text{ AND } (S)\ (\exists P)\ \sim (S \text{ is } P \text{ OR } S \text{ is } \sim P).$$

From Zadeh's "meaning representation" of "words" in Computing With Words, CWW, Peirce's representation is re-expressed as:

$$(S)\ (\exists P)\ (S \text{ is } P \text{ OR } S \text{ is } \sim P)\ (\mu P(S) \in [0,1]) \text{ AND}$$
$$(S)\ (\exists P)\ \sim (S \text{ is } P \text{ OR } S \text{ is } \sim P)\ (1 - \mu P(S)).$$

Here μ is determined by FCM (Bezdek, [2]) or IFC (Celikyilmaz and Turksen, [3]) as will be discussed later in the sequel as:

$$\mu_{ik}^{(t)} = \left[\sum_{j=1}^{c} \left(\frac{d\left(x_k, v_i^{(t-1)}\right)}{d\left(x_k, v_j^{(t-1)}\right)} \right)^{\frac{2}{m-1}} \right]^{-1}.$$

Beyond these essential starting points, it is important to recall that fuzzy system developments were enhanced by t-norms and t-conorms introduced by B. Schweitzer, A. Sklar [10], and certain developments which were proposed by Turksen [13, 14, 15, 16, 17, 18, 19, 20, 21, 22].

2. Introduction

The most commonly known and applied fuzzy system models are "fuzzy rule bases". Such fuzzy rule bases are described by membership functions of the input fuzzy sets that form the left hand sides and the output fuzzy sets that form the right hand sides. This approach was initially proposed by Zadeh [26, 27]. There are two well known and basic variations of this approach with various improvements and enhancements:

(a) The Sugeno-Yasukawa [11] approach where fuzzy sets of both the right and left hand sides are determined either by experts or by fuzzy clustering algorithms such as FCM (Bezdek, [2]).

(b) The Tagaki-Sugeno [12] approach where fuzzy sets of the left hand sides of a fuzzy rule base are determined either by experts or by fuzzy clustering algorithms such as FCM (Bezdek, [2]) and the right hand sides are functions determined by function estimation methods.

But there have been new approaches that propose fuzzy regression function developments in place of fuzzy rule bases. These are:

1) The Hathaway and Bezdek [5] approach, where the determination of a classical regression is enhanced by the introduction of a diagonal membership matrix in the determination of coefficients of a fuzzy regression model where the fuzzy clustering algorithm, FCM (Bezdek [2]), is used to determine the number of such fuzzy regressions required for an effective solution.
2) The Turksen [23] and Celikyilmaz-Turksen [3] approach, where a classical regression is enhanced by the introduction of membership values and their transformations to improve the regression constant, and hence the introduction of fuzzy functions in place of fuzzy rule bases where a fuzzy clustering algorithm such as FCM (Bezdek, [2]) or IFC (Celikyilmaz and Turksen, [3]) is used to determine the number of such fuzzy regressions required for an affective solution.

Next we review these approaches and their essential elements with emphasis on our "Fuzzy Functions", fuzzy regression, models generated by membership values and their transformations.

3. Fuzzy rule base models

Let us first review the fuzzy rule base models in order to identify their unique structures and to point out how they differ from each other.

The most commonly applied fuzzy system models are fuzzy rule bases. Here, we only deal with Multi-Input Single Output (MISO) systems. Generally such 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 which is known as the Zadeh- Fuzzy Rule Base, Z-FRB, can be written as follows:

$$R : \underset{i=1}{\overset{c^*}{ALSO}}(\textbf{IF} \text{ antecedent}_i \ \textbf{THEN} \text{ consequent}_i),$$

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, or IFC. The fuzzy rule base structures determined by alternatives (a) and (b) stated above mainly differ in the representation of the consequents in its structure. If the consequent is represented with fuzzy sets then the fuzzy rule base can be categorized as alternative (a). This is the one initially proposed by Zadeh [27] originally applied by Mamdani, *et al.*, [7], and a modified version is proposed by Sugeno and Yasukawa [11]. Whereas, if the consequents are represented with linear equations of input variables, then the rule base structure is the alternative (b) which we call the Takagi-Sugeno Fuzzy Rule Base 7. These models can be formalized as follows.

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 x in X, can be defined as $X = X_1 \times X_1 \times \cdots \times X_{nv}$, where $X_j \subseteq \Re$ is the domain of the variable x_j. Similarly, the domain of the output variable, y will be denoted as $Y \subseteq \Re$. Then, the i^{th} rule, R_i, and rule base, R, in the structure of Sugeno and Yasukawa [11] can be defined as in (**??**) and (**??**):

$$(3.1) \qquad R_i : \textbf{IF} \ \underset{j=1}{\overset{nv}{AND}} (x_j \in X_j \ \textbf{isr} \ A_{ij}) \ \textbf{THEN} \ y \in Y \ \textbf{isr} \ B_i, \ \forall i = 1, \ \dots, \ c^*$$

$$(3.2) \qquad R : \underset{i=1}{\overset{c^*}{ALSO}}(\textbf{IF} \ \underset{j=1}{\overset{nv}{AND}} (x_j \in X_j \ \textbf{isr} \ A_{ij}) \ \textbf{THEN} \ y \in Y \ \textbf{isr} \ B_i),$$

where A_{ij} is the linguistic label, i.e., fuzzy subset, associated with the j^{th} input variable of the antecedent in the i^{th} rule, R_i, with membership function $\mu_i(x_j) : X_j \to [0, 1]$, and similarly B_i is the consequent linguistic label, i.e., consequent fuzzy subset, of the i^{th} rule with membership function $\mu_i(y) : Y \to [0, 1]$, and c^* is the number of rules in the model. In this structure, the challenges for knowledge representation are: (i) to identify the membership functions of fuzzy sets on the left and right hand sides of the rules and (ii) to identify the most suitable t-norm and t-conorm combinations that represent in a one-to-one correspondence the linguistic "AND" and "OR" for the combination of left hand side fuzzy subsets together with the implication operator, "IMP", that will carry the left hand side membership degree, i.e., the degree of firing, to the right hand side consequent fuzzy subset. As well, one needs to know and be able to apply fuzzy logic to carry out approximate reasoning. It should be recalled that Mamdani [7] applied the Min operator for both "AND and "IMP" which is a very special case whereas the Sugeno and Yasukawa [11] model is more general. It is known that the linguistic "AND" and "OR" operators cannot be represented in a one-to-one correspondence with a particular t-norm and a t-conorm, respectively, as it is shown by Turksen [13, 14, 15, 16, 17, 18, 19, 20, 21, 22].

Hence there must be a selection procedure to determine which t-norm or t-conorm is more suitable for a particular system analysis. Furthermore, for the selected choice of t-norm and t-conorm, one must decide on the use of FDCF and FCCF, Fuzzy Disjunctive and Conjunctive Canonical Forms, which are to be used for the representation of rules and for reasoning with them. However, such models fall into Interval-Valued Type 2 fuzzy systems analyses which are not dealt with in this paper. Finally, one has to carry out defuzzification computations in all fuzzy rule base models. Furthermore the above structure assumes non-interactivity between input variables (Zadeh, [27]). 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. In order to eliminate the non-interactivity assumption, Delgado *et. al.* [4], Babuska *et. al.* [1], and Uncu and Turksen [24] used multi-dimensional Type 1 fuzzy subsets to represent the antecedent part of the rules. In such investigations, generally a multi-dimensional fuzzy clustering technique, e.g., FCM or IFC is implemented to obtain multi-dimensional fuzzy subsets that capture the interactivity (or joint affect) of input variables. Hence, the initial Zadeh's Fuzzy Rule Base (Z-FRB) structure can be expressed as follows:

$$R : \underset{i=1}{\overset{c^*}{ALSO}}(\mathbf{IF}\ x \in X\ \mathbf{isr}\ A_i\ \mathbf{THEN}\ y \in Y\ \mathbf{isr}\ B_i), \tag{3.3}$$

where the multi-dimensional antecedent fuzzy subset of the i^{th} rule is A_i. It should be noted that this multi-dimensional antecedent fuzzy subset determination eliminates the search for the appropriate t-norm for the combination of antecedent fuzzy subsets with "AND".

Thus, in such analyses, the well known variations of Zadeh's (Z-FRB) approach in terms of Sugeno-Yasukawa (SY-FRB) and Takagi-Sugeno (TS-FRB) Fuzzy Rule Base structures are:

$$\begin{aligned} &(SY-FRB)R : \underset{i=1}{\overset{c^*}{ALSO}}(\mathbf{IF}\ x \in X\ \mathbf{isr}\ A_i\ \mathbf{THEN}\ y \in Y\ \mathbf{isr}\ B_i), \\ &(TS-FRB)R : \underset{i=1}{\overset{c^*}{ALSO}}(\mathbf{IF}\ \text{antecedent}_i\ \mathbf{THEN}\ y_i = a_i x^r + b_i), \end{aligned} \tag{3.4}$$

where, antecedent$_i$, $x \in X$ isr A_i, and $a_i = (a_{i,1}, \ldots, a_{i,nv})$ is the regression coefficient vector associated with the i^{th} rule together with b_i which is the scalar associated with the i^{th} rule. For these special cases of Z-FRB, again 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.

4. Fuzzy regression models

In historical order, there are basically two distinct versions of fuzzy regression models. Namely,

(a) The Hathaway and Bezdek [5] approach where the determination of a classical linear regression is enhanced by the introduction of a diagonal membership matrix in the determination of the coefficients of a fuzzy regression model. This work was later extended to a non-linear version by Hoppner and Klawonn [6].

(b) The Turksen [23] and Celikyilmaz-Turksen [3] approach, where a classical regression is enhanced by the introduction of membership values and their transformations to improve the regression constant and hence the introduction of fuzzy functions in place of fuzzy rule bases.

In both of these approaches, a fuzzy clustering algorithm is applied to determine the number of such fuzzy regressions that are required for an effective solution. In case (a) only FCM, Fuzzy C-Means, algorithm (Bezdek, [2]) is applied whereas in case (b) either FCM (Bezdek, [2]) or IFC, Improved Fuzzy Clustering, algorithm (Celikyilmaz and Turksen [3]) is applied in a number of case studies.

4.1. Fuzzy C-regression models, FCRM. Originally the Fuzzy C-Regression Model (FCRM) of Hathaway and Bezdek, [5] was introduced to classify objects into similar groups. FCRM yields simultaneous estimates of parameters for Fuzzy C-Regression models, while fuzzy partitioning a given dataset. It ought to be recalled that FCM is a point-wise clustering algorithm. Furthermore, FCM (Bezdek, [2]) clusters are hypersphere shaped. FCRM determines cluster prototypes as functions instead of geometrical objects. In particular, FCRM determines separate linear patterns, where each pattern can be identified by a linear function. It is to be noted that the FCRM of Hathaway and Bezdek [5] clusters are hyperplane-shaped.

4.1.1. *Differences of FCM and FCRM.* It is well known that the representatives of clusters of FCM are cluster centers, v_i, which are determined by the well known FCM algorithm (Appendix) which can be stated as follows:

$$\begin{aligned}
&\min J(U,V) = \sum_{k=1}^{nd}\sum_{i=1}^{c}(u_{ik})^m(\|x_k - v_i\|)_A,\\
&\text{s.t.} 0 \le u_{ik} \le 1,\ \forall\, i,k,\\
&\sum_{i=1}^{c} u_{ik} = 1,\ \forall k,\\
&0 \le \sum_{k=1}^{nd} u_{ik} \le nd,\ \forall\, i.
\end{aligned} \tag{4.1}$$

Here, J is the 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 $A = C^{-1}$ is the Mahalonobis Norm, etc.

This constraint optimization model can be solved using a well-known method in mathematics, namely the Lagrange Multiplier method, and the model is converted into an unconstraint optimization problem with one objective function. In order to get an equality constraint problem, the primal constraint optimization problem is first converted into an equivalent unconstraint problem with the help of unspecified parameters known as Lagrange Multipliers, λ;

$$\max\ W(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^m d^2(x_k, v_i)_A - \lambda\left(\sum_{i=1}^{c}\mu_{ik} - 1\right).$$

According to the Lagrangian Method, the Lagrangian function must be minimized with respect to primal parameters and maximized with respect to dual parameters. According to the derivative of the Lagrangian function with respect to the original model parameters, U and V should vanish. Hence, by taking the derivative of the objective function with respect to the cluster centers, V and membership values, U, the optimum membership value calculation equation and clusters centers are given by:

$$\mu_{ik}^{(t)} = \left[\sum_{j=1}^{c} \left(\frac{d\left(x_k, v_i^{(t-1)}\right)}{d\left(x_k, v_j^{(t-1)}\right)} \right)^{\frac{2}{m-1}} \right]^{-1},$$

$$v_i^{(t)} = \left(\sum_{k=1}^{n} \left(\mu_{ik}^{(t)}\right)^m x_k \right) \Big/ \sum_{k=1}^{n} \left(\mu_{ik}^{(t)}\right)^m, \ \forall i = 1, \ \ldots, \ c.$$

where $v_i^{(t-1)}$ represents the cluster center vector for cluster i obtained in the $(t-1)^{\text{th}}$ iteration.

It ought to be noted that the μ‘s are highly non-linear transformations of the x‘s, and hence are considered independent of the x‘s. It is important to recall this when we discuss the additive fuzzy functions that were introduced by Turksen [23] and effectively applied by Celikyilmaz-Turksen [3].

Whereas the representatives of clusters in FCRM are hyper-planes, which are represented by:

$$y_i = \beta_i^0 + \beta_i^1 x_1 + \cdots + \beta_i^{nv} x_{nv}$$

where the β_i‘s are the regression coefficients of each function, $i = 1, \ldots, c$.

The FCM algorithm calculates cluster centers by averaging each data vector weighted with their membership values. FCRM calculates cluster representative functions by a weighted least squares regression algorithm such as $y_k = f_i(x_k, \beta_i)$, where $x_k = [x_{1,k}, \ldots, x_{nv,k}]^T \in \Re^{nv}$ denotes the k^{th} data object and $\beta_i \in \Re^{nv}$, $i = 1, \ldots, c$. Performance of these functions is generally measured by:

$$E_{ik}(\beta_i) = (y_k - f_i(x_k, \beta_i))^2.$$

The objective function is to minimize the total error of these approximated functions:

$$E\,(U, \beta_i) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m E_{ik}\,(\beta_i).$$

In FCRM, the μ_{ik}‘s represent how close the extent values predicted by $f_i(x_k, \beta_i)$ are to y_k. It should be recalled that for FCM:

$$\mu_{ik}^{(t)} = \left[\sum_{j=1}^{c} \left(\frac{d\left(x_k, v_i^{(t-1)}\right)}{d\left(x_k, v_j^{(t-1)}\right)} \right)^{\frac{2}{m-1}} \right]^{-1}.$$

Whereas with FCRM one gets:

$$\mu_{ik} = \left[\sum_{j=1}^{c} (E_{ik}/E_{jk})^{\frac{1}{m-1}} \right]^{-1}, \ \forall\, i, j = 1, \ldots, c < n.$$

FCRM is formulated to find hidden structures in a given dataset. Possible extensions of FCRM implement non-linear functions to find hidden patterns.

It is developed with

$$\min : E\,(U, \beta_i) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m E_{ik}\,(\beta_i),$$

where $\beta_i = [X^T U_i X]^{-1} X^T U_i y$,

$$X_i = \begin{bmatrix} x_{i,1}^T \\ x_{i,2}^T \\ \vdots \\ x_{i,n}^T \end{bmatrix},\ y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\ U_i = \begin{bmatrix} \mu_{i1} & 0 & \cdots & 0 \\ 0 & \mu_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_{in} \end{bmatrix}.$$

4.1.2. *Non-linear fuzzy regression.* Hoppner and Klawonn [6] combined the algorithms FCM of Bezdek, [2] and the FCRM of Hathaway and Bezdek [5] in one clustering schema, to build a combined clustering structure. Their aim is to eliminate the counterintuitive membership values. They modify the objective function of FCM by combining it with FCRM.

Hence they determine membership values as:

$$\mu_{ik} = \left[\sum_{j=1}^{c} \frac{d_{ik}^2 - \left(\min_{i=1,\ \ldots,\ c} d_{ik}^2\ - \eta\right)}{d_{jk}^2 - \left(\min_{i=1,\ \ldots,\ c} d_{ik}^2\ - \eta\right)} \right]^{-1},\ 0 < \eta,$$

where $\eta > 0$ is a user defined constant. In Hoppner and Klawonn [6], each function is interpreted as a rule in a Takagi-Sugeno [12] model. For this purpose, Hoppner and Klawonn [6] introduced a new combined distance function, which is the combination of both methods as follows:

$$d_{ik}^2 \left((x_k,\ y_k),\left(v_i,\ \widehat{\beta}_i\right)\right) = \|x_k - v_i(x)\|^2 + \left(y_k - \widehat{\beta}_i^T \hat{x}_k\right)^2,$$

where $\widehat{x}$ represents a user defined polynomial, for instance, a two dimensional polynomial can be formed with the following vector: $(x_1, x_2) = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$.

The coefficients are obtained as:

$$\widehat{\beta}_i = \left(\sum_{k=1}^{n} (\mu_{ik})^m (y_k \hat{x}_k) \right) \Big/ \sum_{k=1}^{n} (\mu_{ik})^m \left(\hat{x}_k \hat{x}_k^T\right),\ \forall i = 1, \ldots, c.$$

In the method proposed by Hathaway and Bezdek [5], first fuzzy clusters are determined by the FCM method to define how many fuzzy regressions are to be constructed, i.e., one for each cluster. There is also the method proposed by Hoppner and Klawonn [6], which combines the FCM, Fuzzy C-Means, and FCRM algorithms in one clustering schema, to build a combined clustering structure. Their main goal is to update the FCM fuzzy clustering algorithm so that they can prevent the effect of harmonics by modifying the objective function. It is to be noted that they not only deal with point-wise clustering algorithms such as the "Fuzzy C-Means" (FCM) clustering algorithm, but as well, they also deal with the "Fuzzy 'C-Regression Model" clustering algorithm (FCRM). It is also well-known that Hathaway and Bezdek, [5] proposed to build linear regression models. Whereas one can build non-linear regression models with the Hoppner and Klawonn [6] approach.

4.2. Fuzzy regression models with LSE, FRM-LSE. The generalization of LSE for Fuzzy Regression Models, called *FRM-LSE* for short, requires that a fuzzy clustering algorithm, such as FCM, or IFC be available to determine the interactive (joint) membership values of the input-output variables in each of the fuzzy clusters that can be identified for a given training data set.

Let (X_k, Y_k), $k = 1, \ldots, nd$, be the set of observations in a training data set, such that $X_k = (x_{jk} \mid j = 1, \ldots, nv,\ k = 1, \ldots, nd)$. 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 the FCM or IFC algorithm, where m is the level of fuzziness (in our experiments we usually take $m = 1.4, \ldots, 2.6$) (Ozkan and Turksen, [9]), and c is the number of clusters (in our experiments we usually take $c = 2, \ldots, 10$). At this point, one ought to recall the well known FCM algorithm is stated in (5) above.

With the application of the well known FCM algorithm is stated in (5) above, one determines the optimal pair (m^*, c^*) is determined. One next identifies the cluster centers for $m = m^*$ and $c = 1, \ldots, c^*$ as:

$$(4.2) \qquad v_{\substack{X|Y,j \\ m^*}} = \left(x^c_{1,j},\ x^c_{2,j},\ \ldots, x^c_{nv,j},\ y^c_j\right).$$

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

$$(4.3) \qquad v_{\substack{X,j \\ m^*}} = \left(x^c_{1,j},\ x^c_{2,j},\ \ldots,\ x^c_{nv,j}\right).$$

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 $\mu'_{ik}s$ that are above an α-cut in order to eliminate harmonics generated by FCM as:

$$(4.4) \qquad u_{ik} = \left(\sum_{j=1}^{c} \left(\frac{\|x_k - v_{X,i}\|}{\|x_k - v_{X,j}\|}\right)^{\frac{2}{m-1}}\right)^{-1},\ \mu_{ik} = \{u_{ik} \geq \alpha\},$$

where μ_{ik} denotes the membership value of the k^{th}, $k = 1, \ldots, nd$, in the i^{th} rule, $i = 1, \ldots, c^*$; x_k denotes the k^{th} vector and for all the input variables $j = 1, \ldots, nv$, in the input space. Once again, it must be emphasized that the μ's are highly non-linear transformations of the x's.

(b) Next, we normalize them as:

$$(4.5) \qquad \gamma_{ij}(x_j) = \mu_{ij}(x_j) \Big/ \sum_{i'}^{c} = 1\mu_{i'} j(x_j).$$

These normalized membership values of x_j, $j = 1, \ldots, nd$ in the i^{th} rule, $i = 1, \ldots, c^*$, the $\gamma_{ij}(x_j)'s$, in turn indicate the membership values that improve the regression constant in our proposed scheme of function identification for the representation of the i^{th} cluster. Thus $\Gamma_i = (\gamma_{ij} | i = 1, \ldots, c^*;\ j = 1, \ldots,\ nd)$ are the membership values of the data sample X in the i^{th} cluster, i.e., the i^{th} rule that causes this improvement in the predictive value of the fuzzy regression model.

Hence we next 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. For the simplest form of representation, an example of a possible augmented input matrix would be: $X'_i = [1,\ \Gamma_i,\ X]$,

$$Y_i = \beta_{i0} + \beta_{i1}\Gamma_i + \beta_{i2}X_{ij}.$$

Alternately, one could consider other augmented input matrices such as $X''_i = \left[1,\ \Gamma^2_i,\ X\right]$ or $X'''_i = \left[1,\ \Gamma^2_i, \Gamma^m_i, \exp\left(\Gamma_i\right),\ X\right]$, etc. for various transformations of $\Gamma_i = (\gamma_{ij} \mid i = 1, \ldots,\ c^*;\ j = 1, \ldots,\ nd)$. The choice depends on whether we want 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 systems 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:

$$(4.6) \qquad X'_{ij} = [1,\ \Gamma_i,\ X_{ij}] = \begin{bmatrix} 1 & \gamma_{i1} & x_{i1} \\ \vdots & \vdots & \vdots \\ 1 & \gamma_{ind} & x_{ind} \end{bmatrix}.$$

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 the FCM algorithm. From this point forward, the estimation of the fuzzy functions takes place for each cluster. Different approaches are followed in the estimation of the fuzzy functions using the augmented matrices. Thus the function $Y_i = \beta_{i0} + \beta_{i1}\Gamma_i + \beta_{i2}X_{ij}$, that represents the i^{th} rule, corresponding to the i^{th} interactive (joint) cluster in $(Y_i,\ \Gamma_i,\ X_j)$ space, would be estimated with the FF-LSE approach as follows:

$$(4.7) \qquad \beta_i^* = \left(X_{ij}'^T X_{ij}'\right)^{-1} \left(X_{ij}'^T Y_i\right),$$

where $\beta_i^* = (\beta_{i0}^*,\ \beta_{i1}^*,\ \beta_{i2}^*)$ are the estimates and $X'_{ij} = [1,\ \Gamma_i,\ X_{ij}]$, provides the inverse of the covariance matrix, if $\left(X_{ij}'^T X_{ij}'\right)^{-1}$, exists. Therefore the estimate of Y_i would be obtained as:

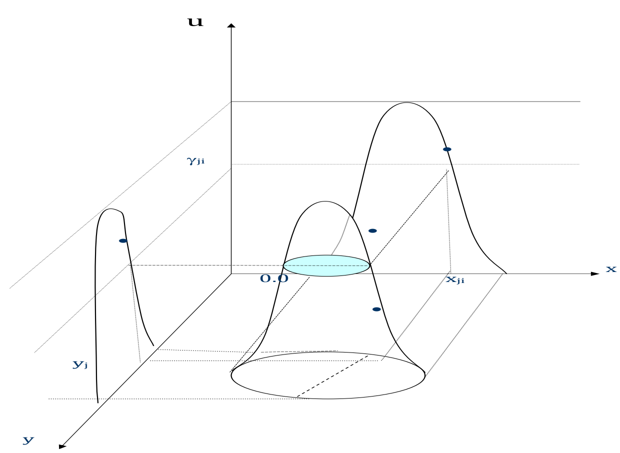
$$(4.8) \qquad Y_i^* = \beta_{i0}^* + \beta_{i1}^*\Gamma_i + \beta_{i2}^* X_{ij}.$$

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

$$(4.9) \qquad Y_i^* = \frac{\sum_{i=1}^{c^*} \gamma_i Y_i^*}{\sum_{i=1}^{c^*} \gamma_i}.$$

Within this 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 the membership values into the augmented input matrix in the space $[U \times X \times Y]$, which can be illustrated with a prototype shown in Figure 1.

Figure 1. A Fuzzy cluster in $[\boldsymbol{U} \times \boldsymbol{X} \times \boldsymbol{Y}]$ space



It ought to be noted for the sake of emphasis that the addition of membership values and or their transformations in fact improves the predictive power of the fuzzy regression equation, for example, in the simplest case:

$$Y_i^* = \beta_{i0}^* + \beta_{i1}^* \Gamma_i + \beta_{i2}^* X_{ij},$$

where $\beta_i^* = (\beta_{i0}^*,\ \beta_{i1}^*,\ \beta_{i2}^*)$ are the estimates and $X'_{ij} = [1,\ \Gamma_i, X_{ij}]$, provided the inverse $\left(X'^T_{ij} X'_{ij}\right)^{-1}$ of the covariance matrix exists. Therefore the estimate of Y_i would be obtained with the addition of the second coefficient associated with Γ in $\beta_i^* = (\beta_{i0}^*,\ \beta_{i1}^*,\ \beta_{i2}^*)$ as a result of the impact of the membership values included in such a regression. It should be once more emphasized that such $X'_{ij} = [1,\ \Gamma_i, X_{ij}]$ improve the initial constant symbolized by "1" of ordinary regression. Once more recall that the Gammas are independent of the input x's, since they are highly non-linear transformations of the input variables!!!

In general, in more complex cases, one may require to include what we call *Enhanced Fuzzy Functions* (Celikyilmaz, and Turksen, [3]) where the membership values represent degrees of *belongingness* with additional *predictors*, which are various transformations of membership values, for example:

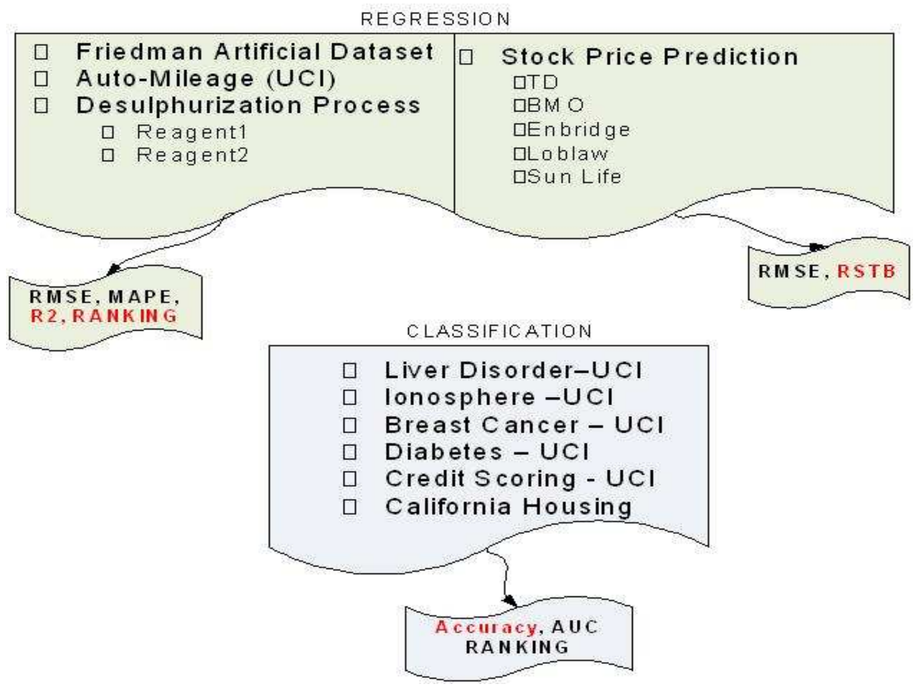
$$h_i(\tau_i,\ \hat{w}_i) = \hat{w}_{0i} + \hat{w}_{1i}\mu_i + \hat{w}_{2i}(1 + \exp(-\mu_{ik}^m)),$$

where the β's are now represented by w's, where an exp transformation of membership values are included as additional information to enhance the constant value of the regression.

5. Experiments – Benchmark datasets

We have experimented with a good number of data sets as shown in Figure 2 below.

Figure 2. Benchmark Data Sets



As well we have conducted various experiments to show which model provides better results as compared to others with respect to various statistical assessments as shown below:

The models we have experimented with are given below in Table 1.

Table 1. Set of Experiments

1	Artificial Neural Networks
2	Support Vector Machines for Regression [Gunn, 1999]
3	ANFIS- Adaptive Neuro-Fuzzy Inference System [Jang,1993]
4	Genetic Fuzzy System [Cordon *et al.*, 2001]
5	DENFIS – Dynamically Evolving Neuro Fuzzy Inference System [Kasabov, 2002]
6	A Type 2 Fuzzy Logic System based on FRB [Uncu, Türkşen, 2007]
7	**Discrete Interval Valued T2FFF, DIT2FF**
8	**Evolutionary Discrete Interval Valued Type 2 IEFF, EDIT2IFF**

In the summary of results, given in Table 2 below, it is shown that models EDIT2IF and DIT2FF provide reasonably "good" results.

Table 2 Summary of Results

Summary of Results

	Best Optimum Methodology	Function Type	Percent Improvement against best optimum benchmark method.
Stock Price Models	EDIT2IFF using Improved Fuzzy Clustering	Linear Regression - LSE	4.46%-28.19% based on CAD$100 investment
Regression Models	EDIT2IFF and DIT2FF Improved Fuzzy Clustering and FCM clustering	LSE and Non-Linear SVM	2%-24.7%
Classification Model	DIT2FF using FCM clustering	Logistic Regression and Non-Linear SVC	3.3%-33.3%

6. Conclusions

We have first reviewed the well known variations of fuzzy rule bases. Then Fuzzy C Regression models of Hathaway and Bezdek as well as of F. Hoppner, F. Klawonn, and next the Fuzzy Regression Models with LSE of Turksen which were further developed by Celikyilmaz and Turksen as alternative models to the fuzzy rule base approaches to fuzzy system modeling. They can be more easily applied by mathematicians and statisticians without knowing the essential mathematical tools of t-norms and t-conorms required for building fuzzy rule base system models.

Furthermore in our applied system investigations, we found that the Fuzzy Regression Models with LSE of Turksen and its further developments by Celikyilmaz and Turksen are better suited for industrial applications and provides better predictions as compared to fuzzy rule based models.

Currently we are working on "Multiplicative Fuzzy Functions" structured as:

$$Y = a + (\mu + \mu_2 + log\mu)x.$$

In contrast to our previous works based on additive structures:

$$Y = a + b\mu + c\mu_2 + dlog\mu + x.$$

That is various transformations of membership values μ are introduced as a multiplier of the input variable x. We expect to publish our results in the near future!!

7. Appendix

The Fuzzy C-means Clustering Algorithm (FCM)

Given: data vectors, $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, number of clusters, $\boldsymbol{c}$, degree of fuzziness, m, and termination constant, ϵ (maximum iteration number in this case). Initialize the partition matrix, U, randomly.

Step 1: Find initial cluster centers using membership values of initial partition matrix as inputs.

Step 2: Start iteration $t = 1 \cdots$ max-iteration value;

Step 2.1. Calculate membership values $\mu_{ik}^{(t)}$ of each input data object k in cluster i, using the membership value calculation, where x_k are input data objects as vectors and $v_i^{(t-1)}$ are cluster centers from the $(t-1)^{\text{th}}$ iteration,

Step 2.2. Calculate the cluster center $v_i^{(t)}$ of each cluster i at iteration t, using the cluster center function, where the inputs are the input data matrix, $\boldsymbol{x}_k$, and the membership values of iteration t, $\mu_{ik}^{(t)}$.

Step 2.3. Stop if the termination condition is satisfied, e.g., $\left|v_i^{(t)} - v_i^{(t-1)}\right| \leq \varepsilon$. Otherwise go to step 1.

References

[1] Babuska, R. and Verbruggen, H. B. *Constructing fuzzy models by product space clustering*, in: H. Hellendoorn and D. Driankov(eds.), Fuzzy Model Identification: Selected Approaches (Springer, Berlin, 1997), 53-90.

[2] Bezdek, J. C. *Pattern recognition with fuzzy objective function algorithms* ISBN-10: 0306406713|ISBN-13: 9780306406713 (1981).

[3] Celikyilmaz, A. and Turksen, I. B. *Enhanced fuzzy system models with improved fuzzy clustering algorithm*, IEEE Trans. Fuzzy Systems **16**, 779–794, 2008.

[4] Delgado, M. Gomez-Skermata, A. F. and Martin, F. *Rapid prototyping of fuzzy models*, in: H. Hellendoorn and D. Driankov (Eds.), Fuzzy Model Identification: Selected Approaches (Springer, Berlin, 1997), 53-90.

[5] Hathaway, R. J. and Bezdek, J. C. *Switching regression models and fuzzy clustering*, IEEE Transactions on Fuzzy Systems **1** (3), 195–204, 1993.

[6] Hoppner, F. and Klawonn, F. *Improved fuzzy partitions for fuzzy regression models*, International Journal of Approximate Reasoning **32**, 85–102, 2003.

[7] Mamdani, E. M. *Application of fuzzy logic to approximate reasoning using linguistic systems* **26**, 1182–1191, 1977.

[8] Mizimoto, M. *Method of fuzzy inference suitable for fuzzy control*, J. Soc. Instrument Control Engineering **58**, 959–963, 1989.

[9] Ozkan, I. and Turksen, I. B. *Upper and lower values for the level of fuzziness in FCM*, Inf. Sci., 5143–5152, 2007.

[10] Schweitzer, B. and Sklar, A. *Probabilistic Metric Spaces* (North-Holland, New York, 1983).

[11] Sugeno, M. and Yasukawa, T. *A fuzzy logic based approach to qualitative modelling*, IEEE Trans on Fuzzy Systems **1** (1), 7–31, 1993.

[12] Takagi, T. and Sugeno, M. *Fuzzy identification of systems and its applications to modelling and control*, IEEE Transactions on Systems, Man and Cybernetics, **SMC-15** (1), 116–132, 1985.

[13] Turksen, I. B. *Interval valued fuzzy sets based on normal forms*, Fuzzy Sets and Systems **20**, 191–210, 1986.

[14] Turksen, I. B. *Four methods of approximate reasoning with interval-valued fuzzy sets*, Int. J. Approximate Reasoning **3**, 121–142, 1989.

[15] Turksen, I. B. *Interval-valued fuzzy sets and compensatory AND*, Fuzzy Sets and Systems **51**, 295–307, 1992.

[16] Turksen, I. B. *Interval valued fuzzy sets and fuzzy connectives*, Interval Computations, 125–142, 1993.

[17] Turksen, I. B. *Interval valued fuzzy sets and fuzzy measures*, Proc. First Int. Conf. of NAFIPS, 317–321, 1994.

[18] Turksen, I. B. *Fuzzy normal forms*, Fuzzy Sets and Systems **69**, 319–346, 1995.

[19] Turksen, I. B. *Type I and "interval-valued" type II fuzzy sets and logics*, in: P. Wang (Ed.), Advances in Fuzzy Set Theory and Technology, 1995.

[20] Turksen, I. B. *Non-specificity and interval-valued fuzzy sets*, Fuzzy Sets and Systems **80**, 87–100, 1996.

[21] Turksen, I. B. *Type I and type II fuzzy system modeling*, Fuzzy Sets and Systems **106**, 11–34, 1999.

[22] Turksen, I. B. *Type 2 representation and reasoning for CWW*, Fuzzy Sets and Systems **127**, 17–36, 2002.

[23] Turksen, I. B. *Fuzzy functions with LSE*, Applied Soft Computing **8**, 1178–1188, 2008.

[24] Uncu, O. and Turksen, I. B. *A novel fuzzy system modeling approach: Multidimensional structure identification and inference*, Proc. Tenth IEEE, International Conference on Fuzzy Systems, (Melbourne, Australia, 2001), 557–562.

[25] Vapnik, N. V. *Statistical Learning Theory* (John Wiley and Sons, New York, 1998).

[26] Zadeh, L. A. *Fuzzy sets*, Information and Control **8**, 338–353, 1965.

[27] Zadeh, L. A. *The concept of a linguistic variable and its application to approximate reasoning*, Information Sciences **8**, 199–249, 1975.