

A Real-Time Analysis of Granular Information: Some Initial Thoughts on a Convex Hull-based Fuzzy Regression Approach

Azizul Azhar Ramli^{1,3}, Witold Pedrycz², Junzo Watada¹, and Nureize Arbaiy^{1,3}

¹Graduate School of Information,
Production and Systems, Waseda
University, 2-7, Hibiikino,
Wakamatsu, Kitakyushu, 808-0135
Japan

²Department of Electrical and
Computer Engineering, University
of Alberta, Edmonton, Alberta,
Canada T6G 2V4 and Systems
Research Institute, Polish Academy
of Sciences, Warsaw, Poland

³Faculty of Computer Science and
Information Technology, Universiti
Tun Hussein Onn Malaysia, Parit
Raja, 86400 Batu Pahat, Johor Darul
Takzim, Malaysia

{azizulazhar@moegi.waseda.jp, pedrycz@ece.ualberta.ca, junzow@osb.att.ne.jp, nureize@uthm.edu.my}

Abstract—Regression models are well known and widely used as one of the important categories of models in system modeling. In this paper, we extend the concept of fuzzy regression in order to handle real-time implementation of data analysis of information granules. An ultimate objective of this study is to develop a hybrid of a genetically-guided clustering algorithm called genetic algorithm-Fuzzy C-Means (GA-FCM) and a convex hull-based fuzzy regression approach being regarded as a potential solution to the formation of information granules. It is anticipated that the setting of Granular Computing will help us reduce the computing time, especially in case of real-time data analysis, as well as an overall computational complexity. We propose an efficient real-time granular fuzzy regression analysis based on the convex hull approach in which a Beneath-Beyond algorithm is employed to design a convex hull. In the proposed design setting, we emphasize a pivotal role of the convex hull approach, which becomes crucial in alleviating limitations of linear programming manifesting in system modeling.

Keywords-convex hull, fuzzy regression, Fuzzy C-Means, genetic algorithm, granular computing

I. INTRODUCTION

Nowadays, we witness a significant growth of interest in Granular Computing (GrC) being regarded as a promising vehicle supporting the design, analysis and processing of information granules [1]. With regard of all processing faculties, information granules are collections of entities (elements), usually originating at the numeric level, that are arranged together due to their similarity, functional adjacency and indistinguishability or alike [1]. Given the similarity function to measure the closeness between the samples these data are clustered into certain granules, categories or classes [2]. The process of forming information granules is referred as information granulation.

GrC has begun to play important roles in bioinformatics, pattern recognition, security, high-performance computing and so on in terms of efficiency, effectiveness, robustness as well as a structural representation of uncertainty [3]. Therefore, the needs for sophisticated intelligent data analysis (IDA) tools are highly justifiable when dealing with this type of information.

Ramli *et al.* proposed a real-time fuzzy regression model incorporating a convex hull method, specifically a Beneath-Beyond algorithm [3]. They have deployed a convex hull approach useful in the realization of data analysis in a dynamic data environment. The developed method exhibits good performance as far as computing time and an overall computation complexity are concerned. Fuzzy C-Means (FCM) clustering algorithm, introduced by Dunn in 1973 and generalized by Bezdek in 1981, becomes one of the commonly used techniques of GrC when it comes to the formation of information granules [4] [5]. There has been a great deal of improvements and extensions of this clustering technique. One can refer here to the genetically-guided clustering algorithm called genetic algorithm-FCM (GA-FCM) and proposed by Hall *et al.* It has been shown that the GA-FCM algorithm can successfully alleviate the difficulties of choosing a suitable initialization of the FCM method [6].

Associated with these two highlighted models (that is fuzzy regression and fuzzy clustering), the main objective of this study is to propose an enhancement of the fuzzy regression analysis for the purpose of analysis of information granules. From the IDA perspective, this research intends to augment the model given originally proposed by Bezdek by including the Ramli *et al.*'s approach. It will be shown that such a hybrid combination is capable of supporting real-time granular based fuzzy regression analysis.

In general, the proposed approach helps perform real-time fuzzy regression analysis in presence of information granules. The proposed approach comprises three main phases. First, we use the GA-FCM clustering to granulate the entire data set into a limited number of chunks—information granules. The second phase consists of constructing convex hull polygons for the already formed information granules. Finally, in the last phase, the selected vertices of the constructed convex hull, which covers all information granules, are used to build a fuzzy regressions model. To illustrate the efficiency and effectiveness of the proposed method, a numeric example is presented.

This study is structured as follows. Section 2 serves as a concise and focused review of the fuzzy regression model, convex hull approach, principles of GrC as well as GA-FCM. Section 3 discusses some essentials of fuzzy linear regression

augmented by the convex hull approach. Section 4 discusses a processing flow of the proposed approach; a real-time granular-based fuzzy regression models. Section 5 is devoted to a numerical experiment. Finally, Section 6 presents some concluding remarks.

II. SOME RELETED PRILIMINARIES

A. Review of Fuzzy Regression Model

In statistical regression, deviations between observed and estimated values are assumed to be due to the random errors. Regression analysis is one of commonly encountered approaches in describing relationships among the analyzed data. The regression models explain dependencies between independent and dependent variables. The variables, which are used to explain the other variable(s) are called explanatory ones [7].

Although conventional regression has been applied to various applications, problems may arise when we encounter vague relationships between input and output variables in which cases there assumptions made for regression models are not valid any longer. This situation becomes a major reason of a lack of relevance of regression models.

Recall that a standard numeric linear regression model comes in the following form

$$y = a_0 + a_1 x_1 + \dots + a_K x_K. \quad (1)$$

As an interesting and useful extension, Tanaka *et al.* [7] introduced an enhancement of the regression model by accommodating fuzzy sets thus giving rise to the term of fuzzy regression or possibilistic regression. The models of this category reflect the fuzzy set based nature of relationships between the dependent and independent variables. The upper and lower regression boundaries in the fuzzy regression are used to quantify the fuzzy distribution of the output values.

As an alternative to the fuzzy specification, an inexact relationship among those dependent and independent variables can be represented via fuzzy linear regression expressed in the following form

$$\tilde{Y} = \tilde{A}_0 X_0 + \tilde{A}_1 X_1 + \dots + \tilde{A}_K X_K = \tilde{\mathbf{A}} \mathbf{X}^t, \quad (2)$$

where $\mathbf{X} = [X_0, X_1, \dots, X_K]$ is a vector of independent variables; $\tilde{\mathbf{A}} = [\tilde{A}_0, \tilde{A}_1, \dots, \tilde{A}_K]$ is a vector of fuzzy coefficients represented in the form of symmetric triangular fuzzy numbers and denoted by $\tilde{A}_j = (\alpha_j, c_j)$. Their membership function is described in the form:

$$u_{\tilde{A}_j}(a_j) = \begin{cases} 1 - \frac{|\alpha_j - a_j|}{c_j}, & : c_j \neq 0, \alpha_j - c_j \leq a_j \leq \alpha_j + c_j \\ 1, & : c_j = 0, \alpha_j = a_j \\ 0, & : otherwise \end{cases} \quad (3)$$

$$(j = 0, \dots, K)$$

In addition, from the computational perspective, the estimation of the membership functions of the fuzzy parameters of the regression is concerned with a certain problem of linear programming (LP) [8] [9].

Given the notation used above, (2) can be rewritten as

$$\tilde{Y}_i = (\alpha_0, c_0) + (\alpha_1, c_1)X_1 + (\alpha_2, c_2)X_2 + \dots + (\alpha_K, c_K)X_K \quad (4)$$

where α_j and c_j are the central (modal) value and the spread of triangular fuzzy number, respectively.

Related to the above model (4), we assume that the numeric input and output data are available, while the relation between the input and output data is defined by some fuzzy function in which the distribution of the parameter is governed by the membership functions [10].

By making use of the extension principle, the membership function of the fuzzy number $\tilde{Y}_i$ comes in the following form

$$u(Y_i) = \begin{cases} 1 - \frac{|Y_i - \mathbf{a} \mathbf{X}_i^t|}{\mathbf{c} | \mathbf{X}_i^t |}, & : \mathbf{c} | \mathbf{X}_i^t | \neq 0, \mathbf{a} \mathbf{X}_i^t - \mathbf{c} | \mathbf{X}_i^t | \leq Y_i \leq \mathbf{a} \mathbf{X}_i^t + \mathbf{c} | \mathbf{X}_i^t | \\ 1, & : \mathbf{c} | \mathbf{X}_i^t | = 0, Y_i = \mathbf{a} \mathbf{X}_i^t \\ 0, & : otherwise \end{cases} \quad (5)$$

where $\mathbf{c} = [c_0, c_1, \dots, c_K]$, $\mathbf{a} = [\alpha_0, \alpha_1, \dots, \alpha_K]$ and

$$\mathbf{c} | \mathbf{X}_i^t | = \sum_{j=0}^K c_j |x_{ij}|, \quad x_{i0} = 1 : (\forall i = 1, \dots, n).$$

Moreover, when we consider an optimization of some objective function, its minimization reduces the total spread of the fuzzy number, $\tilde{Y}_i$. More specifically we have

$$\text{MIN}_{\mathbf{a}, \mathbf{c}} \sum_{i=1}^n \mathbf{c} | \mathbf{X}_i^t | = \text{MIN}_{\mathbf{a}, \mathbf{c}} \sum_{i=1}^n \sum_{j=0}^K c_j |x_{ij}| \quad (6)$$

along with the constraints which require that each observation Y_i comes with at least h degree of membership to $\tilde{Y}_i$, that is $u(Y_i) \geq h$ ($i = 1, \dots, n$). This requirement can be expressed in an equivalent form

$$1 - \frac{|Y_i - \mathbf{a} \mathbf{X}_i^t|}{\mathbf{c} | \mathbf{X}_i^t |} \geq h, \quad \forall i = 1, \dots, n \quad (7)$$

The above formulation of the problem leads to the following problem of LP

$$\text{MIN}_{\mathbf{a}, \mathbf{c}} \sum_{i=1}^n \sum_{j=0}^K c_j |x_{ij}| \quad (8)$$

$$\begin{aligned}
&\text{subject to } \sum_{j=0}^K \alpha_j x_{ij} + (1-h) \sum_{j=0}^K c_j |x_{ij}| \geq Y_i, \\
&\sum_{j=0}^K \alpha_j x_{ij} - (1-h) \sum_{j=0}^K c_j |x_{ij}| \leq Y_i, \\
&c_j \geq 0, \alpha_j \in \mathfrak{R}, x_{i0} = 1 \quad (0 \leq h < 1; \forall i = 1, \dots, n).
\end{aligned} \quad (9)$$

Additionally, the fuzzy regression model is constructed in such a way so that it “covers” all data points under consideration, cf. [11].

We use the notation $(\tilde{y}_i, \mathbf{x}_i)$, $\tilde{y}_i = (y_i, d_i)$, to denote the fuzzy output data and $\mathbf{x}_i = [x_{i0}, x_{i1}, x_{i2}, \dots, x_{iK}]$ ($i = 1, \dots, n$) to describe the numeric input data. The notation of $(\mathbf{a}, \mathbf{c})$ is used to describe fuzzy coefficients. The development of the fuzzy regression model results in the following LP problem:

$$\min_{\mathbf{a}, \mathbf{c}} \sum_{i=1}^n \mathbf{c} | \mathbf{x}_i^t | \quad (10)$$

$$\begin{aligned}
&\text{subject to } y_i + L^{-1}(h)d_i \leq \mathbf{a}\mathbf{x}_i^t + L^{-1}(h)\mathbf{c} | \mathbf{x}_i^t | \\
&y_i - L^{-1}(h)d_i \geq \mathbf{a}\mathbf{x}_i^t - L^{-1}(h)\mathbf{c} | \mathbf{x}_i^t | \\
&\mathbf{c} \geq 0 \quad (i = 1, \dots, n),
\end{aligned} \quad (11)$$

where $L(u)$ refers to the shape of the membership function of the parameters of the fuzzy regression model while h is a possibility degree expressing to which extent a given datum is “covered” by the model. Based on the above explanation, fuzzy regression becomes the possibilistic model that can be used in the context of possibility theory [12]. Furthermore, it is possible in this formulation of the fuzzy regression model to treat non-fuzzy data with zero width setting the value of, d to 0 as described in the above formulas.

B. Convex Hull Approach

Let us recall that the convex hull of a set S of points while $\text{hull}(S)$ is defined to be a minimum convex set containing S . A point $P \in S$ is an extreme point of S if $P \notin \text{hull}(S - P)$. In general, if S is finite, then $\text{hull}(S)$ are a convex polygon and the extreme points of S are the corners of this polygon. The edges of this polygon are referred to as the edges of the $\text{hull}(S)$. Figure 2 illustrates the convex hull realization.

The concept closely related here is the affine hull of a set S in Euclidean space $\mathfrak{R}^K$. It is the smallest affine set containing S , or equivalently, the intersection of all affine sets containing S . Here, an affine set is defined as the translation of a vector subspace. The affine hull $\text{aff}(S)$ of S is the set of all affine combinations of elements of S , namely,

$$\text{aff}(S) = \left\{ \sum_{j=1}^K \alpha_j x_j \mid x_j \in S, \alpha_j \in \mathfrak{R}, \alpha_j \geq 0, \sum_{j=1}^K \alpha_j = 1 \right\}. \quad (12)$$

The Beneath-Beyond algorithm is developed to incrementally build up the convex hull by keeping track of the current convex hull P_i and using an incidence graph [13].

In a nutshell, the Beneath-Beyond algorithm comprises the following steps:

- Step_1** : Select and sort points along one direction, say x_1 . Let $s = \{p_0, p_1, \dots, p_{n-1}\}$ be input points after sorting. Process the points in increasing order.
- Step_2** : Take the first n points, which define a facet (*dimensional faces*) as the initial hull.
- Step_3** : Let P_i be the point to be added to the hull at the i th stage. Let $p_i = \text{conv}(p_0, p_1, \dots, p_{i-1})$ be the convex hull polytope built so far. This step includes two kinds of hull updates;
 - (a) A pyramidal update is done when $p_i \notin \text{aff}(p_0, p_1, \dots, p_{i-1})$ - when P_i is not on the hyperplane defined by the current hull. A pyramidal update consists of adding a new node representing P_i to the incidence graph and connecting this node to all existing hull vertices by new edges.
 - (b) A non-pyramidal update is done when the above condition is not met, i.e. P_i is in the affine subspace defined by the current convex hull. In this case, faces that are visible from P_i are removed and new facets are created.

C. Review of Granular Computing with Genetic Clustering

As said earlier, GrC is a general computing paradigm that effectively deals with processing information granules. The underlying formalism relies on a way in which information granules are represented; here we may consider set theory, fuzzy sets, rough sets, to name a few of the available alternatives [1] [14].

In this context, FCM arises as a way of formation of information granules represented by fuzzy sets [5]. In essence, the FCM algorithm is developed based on the minimization of the following objective function

$$J_m(U, V) = \sum_{i=1}^c \sum_{k=1}^n (U_{ik})^m D_{ik}^2(\mathbf{v}_i, \mathbf{x}_k) \quad 1 \leq i \leq c; 1 \leq k \leq n \quad (13)$$

where $U \in M_{fcn}$ is a fuzzy partition matrix, $V = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c)$ is a collection of cluster centers (*prototypes*). $\mathbf{v}_i \in \mathfrak{R}^K \forall i$ and $D_{ik}(\mathbf{v}_i, \mathbf{x}_k)$ is a distance between x_k and the i^{th} prototype while m is a certain parameter known as a fuzzification coefficient, $m > 1$.

The FCM optimizes (13) by iteratively updating the prototypes and the partition matrix. More specifically, we choose some values of c, m and ε (termination condition – a small positive constant), then generate a random a fuzzy partition matrix U^0 and set an iteration index $t = 0$. An iterative process is organized as follows. Given the

membership value $\mu_{ik}^{(t)}$, the cluster centers $\mathbf{v}_i^{(t)} (i=1, \dots, c)$ are calculated by

$$\mathbf{v}_i^{(t)} = \frac{\sum_{k=1}^n (\mu_{ik}^{(t)})^m \mathbf{x}_k}{\sum_{k=1}^n (\mu_{ik}^{(t)})^m} \quad \text{for } 1 \leq k \leq n. \quad (14)$$

Given the new cluster centers $\mathbf{v}_i^{(t)}$, we update the membership values of the partition matrix $\mu_{ik}^{(t)}$

$$\mu_{ik}^{t+1} = \left[\sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i^{(t)}\|}{\|\mathbf{x}_k - \mathbf{v}_j^{(t)}\|} \right)^{\frac{2}{m-1}} \right]^{-1} \quad 1 \leq j \leq c. \quad (15)$$

This process terminates when $|U^{(t+1)} - U^{(t)}| \leq \varepsilon$, or a predefined number of iterations have been reached [15].

We focus here on the practice of genetically-guided clustering algorithm, which proposed by Hall *et al.* in 1999 [6]. Based on Hall *et al.*, in any generation, element i of the population, V_i , a $c \times s$ matrix of cluster centers or prototypes. The initial population of size P is built by some random assignment of real numbers to each of the features of the c cluster centers. The initial values are constrained to be in the range of the attribute to which they are assigned (determined empirically from the experimental data sets) but are otherwise random [6].

In addition, as the cluster centers are to be used by the GA, it becomes necessary to reformulate the objective function for optimization purposes. Expression (13) can be given in terms of distances from the prototypes (as done in the FCM algorithm). As long as $D_{jk}(\mathbf{v}_j, \mathbf{x}_k) > 0 \forall j, k, 1 \leq j \leq c; 1 \leq k \leq n$ we can use the following formula

$$U_{ik} = 1 / \sum_{j=1}^c \left(\frac{D_{ik}(\mathbf{v}_i, \mathbf{x}_k)}{D_{jk}(\mathbf{v}_j, \mathbf{x}_k)} \right)^{\frac{2}{m-1}} \quad \text{for } 1 \leq i \leq c; 1 \leq k \leq n \quad (16)$$

in the calculations of the fuzzy partition matrix. Therefore the reformulated version of the FCM objective function comes in the form

$$R_m(V) = \sum_{k=1}^n \left(\sum_{i=1}^c D_{ik}^{\frac{1}{1-m}} \right)^{1-m} \quad \text{for } 1 \leq i \leq c; 1 \leq k \leq n \quad (17)$$

Next we optimize R_m by running a genetically-guided algorithm (GGA). As usual, the important features of the GA-based clustering algorithm relate with the choice of the

selection, crossover as well as mutation operators of the genetic optimization; for details one may refer to [6].

Moreover, the complete process of the GGA algorithm is outlined as follows [6].

-
- GGA_1** : Choose m, c and D_{ik} .
- GGA_2** : Randomly initialize P sets of c cluster centers. Constrain the initial values within the space defined by the data to be clustered.
- GGA_3** : Calculate R_m by using (17) for each population member and apply modified objective function $R'_m(V) = R_m(V) + b \times R_m(V)$, where $b \in [0, c]$ is the number of empty clusters.
- GGA_4** : Convert population members to binary equivalents (using the Gray code).
- GGA_5** : For $i = 1$ to number of generations, Do
- i. Used k -fold tournament selection (default $k = 2$) to select $P/2$ parent pairs for reproduction.
 - ii. Complete a two-point crossover and bitwise mutation for each feature of the parent pairs.
 - iii. Calculate R_m by using (17) for each population member and apply modified objective function $R'_m(V) = R_m(V) + b \times R_m(V)$, where $b \in [0, c]$ is the number of empty clusters.
 - iv. Create a new generation of size P which is selected from the two best members of the previous generation and the best children that are generated by using crossover and mutation.
- GGA_6** : Provide the cluster centers to the terminal population with the smallest R'_m value and report R'_m
-

III. FUZZY LINEAR REGRESSION MODELS WITH CONVEX HULL APPROACH

Ordinarily, the weaknesses of the implementation of the multi-dimensional fuzzy linear regression can be addressed by incorporating the convex hull approach [12]. In the introduced modification, the construction of vertices of the convex hull becomes realized in real-time by using related points (convex points) of the graph. Furthermore, the real-time implementation of the method has to deal with a large number of samples (data). Therefore, each particular analyzed sample stands for a convex point and is possibly selected as a convex hull vertex. Some edges connecting the vertices need to be re-constructed as well.

Recall that the main purpose of fuzzy linear regression is to form the upper and lower bounds of the linear regression model. Both the upper line Y^U and lower line Y^L for fuzzy linear regression are expressed in the form:

$$Y^U = \{A_0 + A_1 x_1 + \dots + A_K x_K\}^U : \{\mathbf{A} \mathbf{x}_i^t\}^U = \mathbf{a} \mathbf{x}_i^t + \mathbf{c} \left| \mathbf{x}_i^t \right| \quad (18)$$

$$Y^L = \{A_0 + A_1x_1 + \dots + A_Kx_K\}^L : \{\mathbf{A}\mathbf{x}_i^t\}^L = \mathbf{a}\mathbf{x}_i^t - \mathbf{c}|\mathbf{x}_i^t| \quad (19)$$

By using (18) and (19), we convert the problem to a general fuzzy regression that is similar to the one shown below:

(i) Evaluation (objective) function

$$\min_{\mathbf{a}, \mathbf{c}} \sum_{i=1}^n \sum_{j=2}^K c_j |P_{ij}| \quad (20)$$

(ii) Constraints

$$P_{i1} \in Y_i \Leftrightarrow \begin{cases} P_{i1} \leq \alpha_0 + c_0 + \sum_{j=2}^K \alpha_j P_{ij} + \sum_{j=2}^K c_j |P_{ij}| \\ P_{i1} \geq \alpha_0 - c_0 + \sum_{j=2}^K \alpha_j P_{ij} - \sum_{j=2}^K c_j |P_{ij}| \end{cases} \quad (i = 1, \dots, n) \quad (21)$$

The above expression can be further rewritten as follows:

$$\begin{aligned} Y^U &= \{Y_i^U \mid i = 1, \dots, n\} \\ Y^L &= \{Y_i^L \mid i = 1, \dots, n\}. \end{aligned} \quad (22)$$

We also arrive at the following simple relationships for P_{i1}

$$P_{i1} \leq Y_i^U, \quad P_{i1} \geq Y_i^L \quad (i = 1, \dots, n). \quad (23)$$

It is well known that any discrete topology is a topology which is formed by a collection of subsets of a topological space X and the discrete metric ρ on X is defined as

$$\rho(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases} \quad (24)$$

for any $x, y \in X$. In this case, (X, ρ) is called a discrete metric space or a space of isolated points. According to the definition of discrete topology, expression (23) is rewritten as follows:

$$\begin{aligned} S(Y^U) &= \sum_{j=1}^K \{Y_j P_{ij}\}^U \geq 0 \\ S(Y^L) &= \sum_{j=1}^K \{Y_j P_{ij}\}^L \leq 0 \end{aligned} \quad (25)$$

where we assume that $P_{i1} = 1$.

This formula corresponds to the definition of the support hyperplane. Under the consideration of the range of

$$S \cap P \neq \emptyset \text{ and } P \subset S^+ \text{ or } P \subset S^-, \quad (26)$$

the following relation is valid:

$$\bigcap S(Y^U) = \bigcap S(Y^L). \quad (27)$$

This is explained by the fact that regression formula Y^U and Y^L are formed by vertices of a convex hull. Therefore, it is apparent that the constructed convex hull polygon or more specifically, its vertices clearly define the as discussed constraints of fuzzy mathematical programming, which becomes more reliable as well as significant for the subsequent processes.

Recall that the convex hull of a set S of points while $\text{hull}(S)$ is defined to be a minimum convex set containing S . A point $P \in S$ is an extreme point of S if $P \notin \text{hull}(S - P)$. Hence P denotes the set of points (input samples) and P_C is the set of vertices of the convex hull where $P_C \in P$. Therefore, the convex hull has to satisfy the following relationship:

$$\text{conv}(P) = \text{conv}(P_C) \quad (28)$$

Let us introduce the following set

$$P_C = \{x_{Cl} \in \mathbb{R}^K \mid l = 1, \dots, m\} \subseteq P \quad (29)$$

where m is the number of vertices of the convex hull. Plugging this relationship into (21), we arrive at the following constraints:

$$P_{i1} \in Y_i \Leftrightarrow \begin{cases} P_{i1} \leq \alpha_0 + c_0 + \sum_{j=2}^K \alpha_j P_{ij} + \sum_{j=2}^K c_j |P_{ij}| \\ P_{i1} \geq \alpha_0 - c_0 + \sum_{j=2}^K \alpha_j P_{ij} - \sum_{j=2}^K c_j |P_{ij}| \end{cases} \quad (i = 1, \dots, m). \quad (30)$$

By using (30), the constraints of the LP of the fuzzy linear regression can be written down in the following manner:

$$y_i \in Y_i \Leftrightarrow \begin{cases} y_i \leq \mathbf{a}\mathbf{x}_i^t + \mathbf{c}|\mathbf{x}_i^t| \\ y_i \geq \mathbf{a}\mathbf{x}_i^t - \mathbf{c}|\mathbf{x}_i^t| \end{cases} \quad (i = 1, \dots, m). \quad (31)$$

Moreover, in order to form a suitable regression model based on the constructed convex hull, the connected vertex points are used as the constraints in the LP formulation of the fuzzy linear regression. Considering this process, the use of

the limited number of selected vertices contributes to the minimized computing complexity associated with the model [1].

IV. A REAL-TIME GRANULAR BASED FUZZY REGRESSION MODELS WITH A CONVEX HULL IMPLEMENTATION

In what follows, we present a hybrid combination of a fuzzy linear regression analysis based on convex hull algorithm, in particular the Beneath-Beyond algorithm to facilitate a real-time fuzzy regression analysis of granular data implementation. The detailed flow of processing is shown in Figure 1.

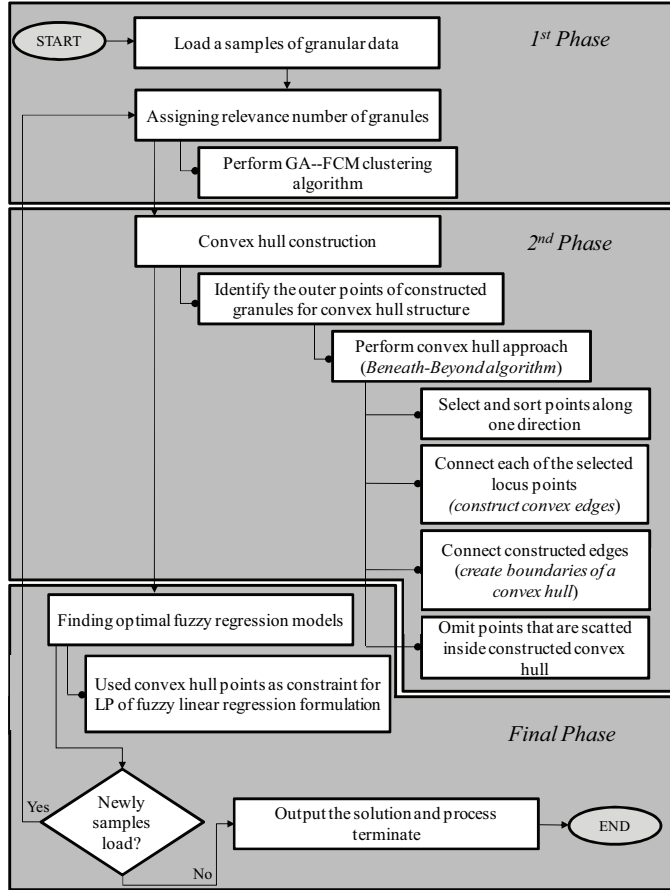


Figure 1. A general flow of processing of the proposed approach.

During the first phase we perform an assigning of appropriate number of granules augmented by GA-FCM clustering algorithm. Then, at the next phase, a construction of convex hull polygon is realized by selecting the outer points among the assigned granules. Therefore, the connection between these selected points (vertices) were became a boundary of the entire analyzed data samples as well as obtained information granules. As a result, with this constructed convex hull, we create the optimal regression lines as realized in the final phase, see Figure 1. In addition, we complete some iterations of the overall procedure considering that more data become available. Say, new samples are provided within a certain time interval, e.g., they could be arriving every 5 or 10 seconds.

In light of the comments made above, it becomes apparent that the quality of granular based fuzzy regression model can be improved by the hybrid combination of fuzzy regression approach with the convex hull approach. The efficiency refers to the computing time as well as the overall computational complexity.

We do not have to consider the complete feature vectors for building regression models; just we utilize the selected vertices which are used for the construction of the convex hull. Therefore, this situation will lead to the decrease of computation load. On the other hand, related to the computational complexity factor for the subsequent iteration, it will only consider the newly added samples of data together with the selected vertices of the previous convex hull. For that reason, this computing scenario will reduce the computational complexity because of the lower number of the feature vectors used for the subsequent processing of regression models [3].

V. A NUMERICAL EXAMPLE

We present a simple numerical example, which shows the efficiency of the proposed approach in the implementation of real-time granular based fuzzy regression. Below, shows an illustration of real-time data analysis which involve dynamic record/database, see Figure 2. For instant, each of iteration might be having a same amount of newly arrived data. As early mentioned, the size of data were increase along with time spending. We can see here, the volume of data becomes enlarge as regard to the time increased.

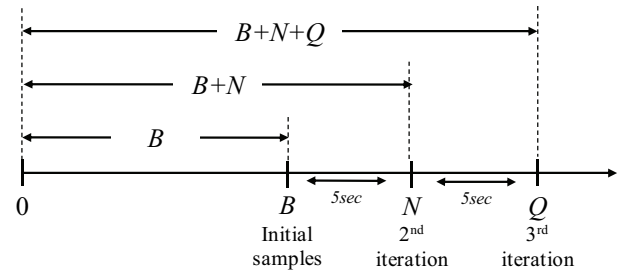


Figure 2. An Illustration of an increasing record/database along with time consumption.

Based on the above figure, we assume that an initial group of samples (for 1st iteration) consists of 100 data of the well-known *Iris* data set [11]. Next, considering a distribution of these data, we construct a convex hull polygon, which becomes the boundary of the overall analyzed samples. Among 100 obtained loci points (clustered feature vectors), only 11 points were selected as convex hull vertices, see Figure 3. In addition, these selected vertices are located as the outside points of the constructed clusters. By solving the associated LP problem that considered these selected vertices as a part of the constraint portion standing in the problem, we obtained the optimal regression coefficients, see Table I. In addition, we selected $h = 0.05$ to express goodness of fit or compatibility of data and the regression model.

TABLE I. THE CENTER AND WIDTH VALUES OF INITIAL GROUP OF DATA SAMPLES

Variables	Center, m_i	Width, c_i
Intercept	2.071	0.163
C1	0.612	0.096
C2	0.639	0.075
C3	-0.412	0.000

Figure 2 shows the obtained feature vectors.

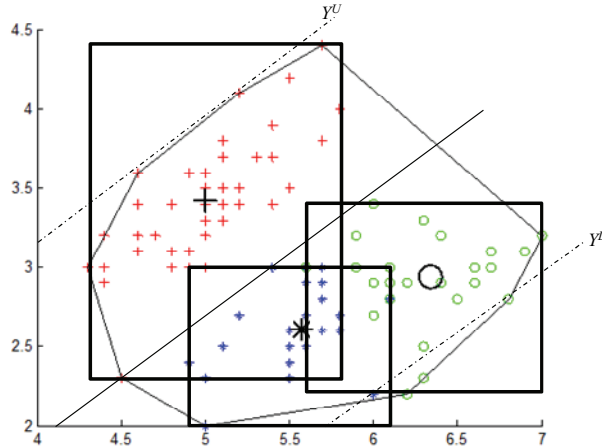


Figure 3. An Illustration of obtained clustered feature vectors for the initial group of data samples.

To deal with a real-time scenario, we added a group of samples taken from the same data set, which consists of 50 patterns. In this case, we assume that a same as previous parameter setting has been set up and an iteration process has been completed. For this 2nd iteration, only 9 points were considered as convex hull vertices.

Finally, Table II includes the parameters as well as the values of the fuzzy regression which have been obtained and Figure 4 shows the of clustered feature vectors.

TABLE II. CENTER AND WIDTH VALUES OF INITIAL GROUP TOGETHER WITH NEWLY ADDED DATA SAMPLES

Variables	Center, m_i	Width, c_i
Intercept	1.856	0.173
C1	0.651	0.102
C2	0.709	0.095
C3	-0.556	0.000

Let us highlight some essential features related to the time analysis. In general, an increasing number of potential vertices present in the subsequent iterations may not directly influence the required time analysis because some of vertices, which have been previously selected, might be reused in the consecutive iteration. Therefore, this situation might reduce required computing time. Furthermore, due to the decrease in the number of considered vertices, this effect will influence (*reduce*) the computational complexity.

On the other hands, as a supporting evidence for evaluating the performance as well as efficiency of this proposed

approach, we use the same set of data samples/parameters and each of experiment was run five times for some other regression approaches; LP approach and fuzzy least-squares regression (FLSR) [16]. Each run presents samples of data sequentially in different order depending on their own procedure. Table III shows the performance of the proposed approach and the remaining regression models in time consumption point of view.

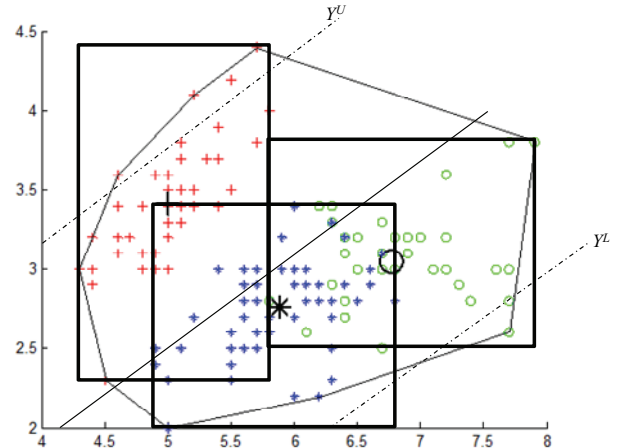


Figure 4. Illustration of obtained clustered feature vectors for initial group together with newly added data samples.

TABLE III. PERFORMANCE ANALYSIS-TIME CONSUMPTION FACTOR: DETAILS

Model	Iteration	Points Total (Considered Vertices)	Time Used (seconds)
LP approach	1 st iteration	100 (100)	0.58
	2 nd iteration	150 (150)	1.22
FLSR approach	1 st iteration	100 (100)	0.47
	2 nd iteration	150 (150)	1.04
Proposed approach	1 st iteration	100 (11)	0.33
	2 nd iteration	150 (9)	0.39

Based on above table, we can summarized that, our proposed approach was utilized the minimal time consumption. In addition, noted that, those, LP and FLSR regression approach was not appropriate for dealing with dynamic data analysis. Therefore, these results illustrate the potential of our proposed approach to solve real-time fuzzy regression problem especially for granular computing fields that demand incremental adaptability.

VI. CONCLUDING REMARKS

In this study, we have developed an enhancement of the IDA tool of fuzzy regression in case of information granules. The method first constructs a limited number of information granules and afterwards those granules are processed by running the convex hull-based fuzzy regression. In this way, we have successfully developed a new idea of real-time granular based fuzzy regression models as a modeling alternative to deal with real-world regression problems.

It is shown that information granules are formed as a result of running the genetic version of the FCM called GA-FCM. The second phase of the overall method is completed depending upon the outer plots of the clusters. The sequential flow of processing was carried out to deal with dynamically increasing data set.

Based on the experimental developments, one could note that this approach becomes a suitable design alternative especially when solving real-time fuzzy regression problems with information granules. It works efficiently for real-time data analysis given the reduced processing time as well as the associated computational complexity.

This proposed approach can be applied to real-time fuzzy regression problems in large-scale systems presented in real-world scenarios. In addition, each of selected approach (*GA-FCM and convex hull-based fuzzy regression*) has their own features to be faced with dynamically changes of samples volume within a certain time interval. Therefore, as a result, this extended enhancement or hybrid combination will provide an efficient platform for regression purposes which require only plenty of time as well as reduce computational complexity with lower computing overhead. Although in this paper we dealt with small data sets (*and this was done for illustrative purposes*), it is worth noting that method scales up quite easily.

In further studies, we plan to expand the proposed approach by incorporating some other technologies of soft computing and swarm intelligence.

ACKNOWLEDGMENT

The first author was supported by Ministry of Higher Education Malaysia (MOHE) under *Skim Latihan Akademik IPTA (SLAI-UTHM)* scholarship program at Graduate School of Information, Production and Systems (IPS), Waseda University, Fukuoka, Japan.

REFERENCES

- [1] A. Bargiela and W. Pedrycz, "Granular computing: An introduction," *Kluwer Academic Publishers*, 2003.
- [2] D. Shifei, X. Li, Z. Hong and Z. Liwen "Research and progress of cluster algorithms based on granular computing," *International Journal of Digital Content Technology and its Applications*, Volume 4, Number 5, 2010, pp. 96-104.
- [3] A. A. Ramli, J. Watada and W. Pedrycz, "Real-time fuzzy regression analysis: A convex hull approach," *European Journal of Operational Research*, Volume 210, Issue 3, 1 May 2011, pp. 606-617.
- [4] A. Bargiela and W. Pedrycz, "Toward a theory of granular computing for humancentered information processing," *IEEE Transactions on Fuzzy Systems*, Volume 16, Issue 16, 2008, pp. 320-330.
- [5] B. Chen, P. C. Tai, R. Harrison and Y. Pan, "FIK model: Novel efficient granular computing model for protein sequence motifs and structure information discovery," *Sixth IEEE International Symposium on Bioinformatics and BioEngineering (BIBE 2006)*, 16-18 October 2006, Arlington, Virginia, USA, pp. 20-26.
- [6] L. O. Hall, I. B. Ozyurt and J. C. Bezdek, "Clustering with a genetically optimized approach," *IEEE Transactions on Evolutionary Computation*, Volume 3, Issue 2, 1999, pp. 103-112.
- [7] J. Watada and W. Pedrycz, "A possibilistic regression approach to acquisition of linguistic rules," in: W. Pedrycz, A. Skowron and V. Kreinovich (eds.), *Handbook on Granular Commutation*, John Wiley and Sons Ltd, Chapter 32, 2008, pp. 719-740.
- [8] J. Watada, Y. Toyoura and S. G. Hwang, "Convex hull approach to fuzzy regression analysis and its application to oral age model," *International Fuzzy System Association World Congress and 20th North American Fuzzy Information Processing Society International Conference 2001*, Joint 9th, Canada, 2001, pp. 867-871.
- [9] H. Tanaka, S. Uejima and K. Asai, "Linear regression analysis with fuzzy model," *IEEE Transactions Systems, Man and Cybernetics*, Volume 12, No. 6, 1982, pp. 903-907.
- [10] H-F. Wang and R-C. Tsaor, "Insight of a possibilistic regression model," *Fuzzy Sets and Systems*, Volume 112, Issues 3, 2000, pp. 355-369.
- [11] A. Frank and A. Asuncion, *UCI Machine Learning Repository* [<http://archive.ics.uci.edu/ml/>]. Irvine, CA: University of California, School of Information and Computer Science, 2010.
- [12] A. A. Ramli, J. Watada and W. Pedrycz. "Real-time fuzzy switching regression analysis: a convex hull approach," *11th International Conference on Information Integration and Web-based Applications and Services (iiWAS2009)*. Kuala Lumpur, Malaysia, 14-16 December 2009, pp. 284-291.
- [13] B. Barber B., D. P. Dobki, H. Hupdanpaa "The quickhull algorithm for convex hull," *ACM Transaction on Mathematical Software*, Vol. 22, No. 4, 1996, pp. 469-483.
- [14] B. Apolloni, D. Iannizzi, D. Malchiodi and W. Pedrycz, "Granular regression," *Neural Nets. 16th Italian Workshop on Neural Nets, WIRN 2005 and International Workshop on Natural and Artificial Immune Systems, (NAIS 2005)*, Vietri sul Mare, Italy, 2005, Lecture Notes in Computer Science, Volume 3931, 2006, pp. 147-156.
- [15] S. Nascimento, B. Mirkin and F. Moura-Pires, "A fuzzy clustering model of data and fuzzy c-means," *IEEE Conference on Fuzzy Systems (FUZZ-IEEE2000)*, San Antonio, Texas, USA, 2000, pp. 302-307.
- [16] C. Kao and C. Chyu, "Least-squares estimates in fuzzy regression analysis," *European Journal of Operational Research*, Vol. 148, Issue 2, 2003, pp. 426-435.