# Fuzzy Regression for Weight Information Extraction in Fuzzy Environment

**Nureize A.**
*Universiti Tun Hussein Onn Malaysia (UTHM),*
*nureize@uthm.edu.my*

## ABSTRACT

Extracting new information is one of the knowledge management ability yet difficult in environment which contains fuzzy information. For example, measuring a product's quality is an important task which normally requires visual inspection basis, and this grading is subject to expert knowledge and interpretation. Apparently, fuzzy information is inherently included in this evaluation and an appropriate tool is necessary to manage such information. Thus, to assist the uncertain situation, in this paper, a fuzzy regression is introduced to improve the extraction of attribute'sweight in a multi-attribute decision making. The proposed model will manage the linguistic assessment provided by evaluators in order to compute collective assessments about the product samples. The proposed model is applied to milled rice grading, as the quality inspection process requires a method to ensure product quality. We include simulation results and highlight the advantage of the proposed method in handling the existence of fuzzy information and managing the information which comes from such uncertain environment.

**Keywords**:Fuzzy regression analysis, multi-attribute, attributes weighting, decision making

## 1 INTRODUCTION

The ability to generate mass information at present technology is apparently caused the improvement to the information and knowledge management strategy. In present work, knowledge has been obtainedfrom numerous different sources such as published literature, human specialists, users and existingmodels (Nureize, 2004).The creation and diffusion of knowledge havebecome ever more important factors in any type of competitiveness (Dalkir, 2005). Progressively,knowledge is being extremely important and powerful to sufficiently extracted and organized. From small scale of data to the massive amount of data, extracting and managing information is crucial to obtain the benefits from the extracted knowledge. From last decade, many tools and methodology has been introduced and improved to handle the amount of data and the strategies to obtain the knowledge.

One of the available technique which has been widely optimized the usage is known as multi-attribute decision making. Determining an attribute's weight (i.e. in multi-attribute decision making) is sometimes difficult if relevant data are either unavailable or difficult to obtain, and thus requires additional extraction task. Additionally the assignments of attribute weights may vary from one decision maker to another. Therefore, an appropriate method is required for determining these weights, as these decisions are crucial to the model's performance. There exist methods which can be used to generate the attribute weight to alleviate the difficulties. For example, a regression analysis is one of the possible methods used to estimate the weights of the model (Nureize and Watada, 2010).

Regression analysis is a statistical technique for modeling the relationship between variables (Montgomery *et al.*, 2012). This technique is supported by effective statistical analysis dealing with numeric precise data. Regression analysis is one of the techniques uses in statistical data for coefficient estimation. This technique is supported by effective statistical analysis dealing with precise crisp data. The importance of this regression approach is to investigate and model the relationship between variables. This technique is widely used in forecasting and prediction and had been used over a decade's (Huffle and Vandewall, 1987; Yang and Ko, 1997; Chertow *et al.*, 1997; Montgomery *et al.*, 2012). Thus, the regression analysis have a significant role in finding relationship between variables and apparently useful for various application.

A regression method analyzes statistical data to estimate the model coefficients in developing effective models. The conventional mathematical programming problem uses numerical deterministic values to these coefficients. In contrary, it is more realistic to take the estimated values of the coefficients as imprecise values rather than precise ones. In real life application, there exists uncertainty fuzzy data in which subjective human estimation play a main role. As a result, a fuzzy regression models was introduced to cope with the fuzzy uncertainty input-output data (Tanaka *et al.*, 1982).

In practical systems, probabilistic or/and vague situations include uncertain information such as predictions of future profits and incomplete historical data. Therefore, the mathematical programming models should be able to handle the above problems. That is, the above situations should be explicitly considered in the decision making process and information management strategies. For that reason, the fuzzy regression model is introduced to improve information extraction with the existence of the fuzziness in the data used for the approximation (Nureize and Watada, 2010). The property of fuzzy regression model is used to allow for the co-existence of fuzziness in the data. Thus, the information of weight value which is deduced by the fuzzy regression model is useful to further build a decision making model.

The remainder of this paper is organized as follows: related research studies are reviewed briefly in Section 2. Section 3 describes the fuzzy evaluation model. The fuzzy weights in the model are assessed by fuzzy regression analysis in Section 4. Section 5 describes the numerical experimentation and its findings. Section 6 concludes this paper with some additional remarks.

## II. RELATED WORKS

Fuzzy regression is a regression analysis on fuzzy data in dealing with the fuzziness environment (Yang and Ko, 1997; Chang and Ayyub, 2001). The first linear of regression analysis with fuzzy model was proposed by Tanaka *et al.*, (1982). This method was introduced as to cope with the imprecise data coming from fuzzy environments where human subjective estimates play a main role. Fuzzy regression has been studied and applied widely in the fuzzy environment. In knowledge management domain of study, fuzzy theories has played its important role and successfully applied (e.g., Khoshsima *et al.*, 2004; Zhai *et al.*, 2008).

The study of fuzzy regression technique is continues by improving the fuzzy regression (Tanaka *et al.* 1988) with fuzzy input-output data using shape-preserving fuzzy arithmetic operation (Song et al., 2005). Moreover, new method for computation of fuzzy regression was proposed where the approach is depends on the entropy's properties to rectify previous problems in fuzzy linear regression with crisp input and fuzzy output (Pasha et al., 2007; Höppner and Klawon, 2003). The fuzzy regression also applied as to solve problem in agricultural industry where it used to decide the fuzzy weight based on the criteria of oil palm fruit (Nureize and Watada, 2008). Meanwhile, different approach has also been proposed for computation of fuzzy linear regression based on the goal programming technique to estimate upper and lower fuzzy band with fuzzy output and crisp inputs (Abdollahzadeh *et al.*, 2010).

Classical fuzzy regression is written as follows:

$$Y = [Y_j] = [A_1 x_{j1} + A_2 x_{j2} + \cdots + A_n x_{jn}] = \mathbf{A}\mathbf{x}_j{}' \qquad x_{j1} = 1; j = 1,2,\cdots n \tag{1}$$

where regression coefficient $A_i$ is a triangular-shaped fuzzy number $A_i = \langle a_i, h_i \rangle$ with centre $a_i$ and spread $h_i$. In Equation (1), $\mathbf{x}_j$ is a value vector of all criteria for the $j^{th}$ sample.

Referring to the extension principle, Equation (1) is rewritten as follows:

$$Y_j = \mathbf{A}\mathbf{x}_j{}' = \langle \mathbf{a}\mathbf{x}_j{}', \mathbf{h} \,|\, \mathbf{x}_j \,|' \rangle \tag{2}$$

where $|\mathbf{x}_j| = (|x_{j1}|, |x_{j2}|, \cdots, |x_{jK}|)$. The output of the fuzzy regression (1) results in fuzzy number, whose coefficients are fuzzy numbers.

The regression model with fuzzy coefficients can be described using the lower boundary $\mathbf{a}\mathbf{x}_j{}' - \mathbf{h}\,|\,\mathbf{x}_j\,|'$, centre $\mathbf{a}\mathbf{x}_j{}'$ and upper boundary $\mathbf{a}\mathbf{x}_j{}' + \mathbf{h}\,|\,\mathbf{x}_j\,|'$. A sample $(y_j, \mathbf{x}_j)$, $j = 1,2,\cdots,n$ is defined for the total evaluation with centre $y_j$, spread $d_j$ as a fuzzy number $y_j = \langle y_j, d_j \rangle$. The inclusion relation between the model and the samples should be written as Equation 2.3:

$$\begin{aligned} y_j + L^{-1}(\alpha) d_j &\le \mathbf{a}\mathbf{x}_j{}' + L^{-1}(\alpha)\mathbf{h}\,|\,\mathbf{x}_j\,|' \\ y_j - L^{-1}(\alpha) d_j &\ge \mathbf{a}\mathbf{x}_j{}' - L^{-1}(\alpha)\mathbf{h}\,|\,\mathbf{x}_j\,|' \end{aligned} \tag{3}$$

In other words, the fuzzy regression model is built to contain all samples in the model. This problem results in a linear program.

Using the notations of observed data $(y_j, \mathbf{x}_j)$, $y_j = \langle y_j, d_j \rangle$, $\mathbf{x}_j = [x_{j1}, x_{j2}, \cdots, x_{jK}]$ for $j = 1, 2, \cdots, n$ and fuzzy coefficients $\mathbf{A}_i = \langle \mathbf{a}_i, \mathbf{h}_i \rangle$ for $i = 1, 2, \cdots, K$, the regression model can be mathematically written as the following linear programming problem:

$$\begin{aligned} \min_{\mathbf{a},\mathbf{h}} \quad & \sum_{j=1}^{n} \mathbf{h} |\mathbf{x}_j|^t \\ \text{subject to } & y_j + L^{-1}(\alpha) d_j \le \mathbf{a}\mathbf{x}_j^t + L^{-1}(\alpha)\mathbf{h}|\mathbf{x}_j|^t, \; j = 1, 2, \cdots, n, \\ & y_j - L^{-1}(\alpha) d_j \ge \mathbf{a}\mathbf{x}_j^t - L^{-1}(\alpha)\mathbf{h}|\mathbf{x}_j|^t, \; j = 1, 2, \cdots, n, \\ & \mathbf{h} \ge 0 \end{aligned} \quad (4)$$

A fuzzy regression is obtained by solving the linear programming problem in Equation (4). The fuzzy regression model is also capable to treat non-fuzzy data by setting the spread to 0 in the above equations.

## III. FUZZY REGRESSION EXTRACTIONMODEL

The fuzzy regression extraction model (FREM) uses ratio between weights for scoring task. By means. a straightforward judgment is used to compare the attributewith each other. Triangular fuzzy numbers instead of crisp numbers are used to describe the fuzzy importance level. A triangular fuzzy number is denoted by $A = (a, h)$ using central value $a$ width $h$. Table I shows the intensity of a compliance scoring scale for a crisp number and a fuzzy number.

A combination of crisp and fuzzy numbers is used based on the appropriateness with the attribute of the problem. and is assigned to the alternatives to measure their performance against each criterion. The mixture of crisp and fuzzy numbers can give flexibility and extension to the evaluation process where a suitable judgment scale can be made that corresponds to the attribute.

In this study, the general information of weight extraction is enhanced using a fuzzy regression method. This decision making process consists of three stages listed in the following:

### *A. Review Related Reference and Information Acquisition*

The initial step in the decision framework is to review related references to accumulate the key pieces of knowledge in the study domain. With the advancement of technology, greater information and knowledge have been properly documented and published digitally. These documents can be used as a reference. On the other hand. expert interviews and brainstorming can also be arranged in order to gain additional insight and validate the findings from published references.

### *B. Weight Extraction Using Fuzzy Regression*

Fuzzy regression analysis was used to model an expert evaluation structure. Scoring for each respective attributes is obtained by the human examiner using information in Table I. Then fuzzy regression model (4) is executed to extract the weight values.

### *C. Decision Making and Analysis*

The obtained weight information values are analyzed to determine the significance or contribution to the total assessment of the respective problem in study.

A steps of solution is constructed at which aimed to assist the information extraction (weight) and to further proceed with decision making such as analysis of the most contributing attributes and ranking the sample.

TABLE I
INTENSITY OF COMPLIANCESCALE

| Intensity of Importance | | | Definition |
|---|---|---|---|
| Crisp value | Fuzzy value | | |
| | Notation | Membership Function | |
| 1 | $\tilde{1}$ | (1,1) | Serious noncompliance |
| 2 | $\tilde{2}$ | (2,1) | Serious noncompliance to moderately noncompliance |
| 3 | $\tilde{3}$ | (3,1) | Moderate noncompliance |
| 4 | $\tilde{4}$ | (4,1) | Moderate noncompliance to compliance |
| 5 | $\tilde{5}$ | (5,1) | Comply |
| 6 | $\tilde{6}$ | (6,1) | Comply to exceed compliance |
| 7 | $\tilde{7}$ | (7,1) | Exceed compliance |
| 8 | $\tilde{8}$ | (8,1) | Exceed to substantially exceed compliance |
| 9 | $\tilde{9}$ | (9,1) | Substantially exceed compliance |

Read as:
"The product sample is ____________ with the standard

## VI. ILLUSTRATIVE EXAMPLE AND DISCUSSIONS

This section shows an example of weight extraction using fuzzy regression FREM. Milled rice evaluation process is used as a case study in this experiment. Since rice is the essential food commodity, continuously rice supply is needed. Rice production in this country is not being able to meet this demand even though the commodities depend on rice food energy (Fahmi *et al.*,2013; Daño and Samonte, 2005). The lower quantity and quality of rice production cause the industry of rice production fail to meet the demand. Due to this reason, many researcher and institution of paddy in Malaysia contribute to help enhancing the competitiveness and profitability in agriculture and forestry. Another work on artificial intelligent technique for rice grading can be found elsewhere (Silva, and Sonnadara. 2013, Aulakh and Banga, 2012; Kaur and Verma, 2013).

Four attribute were considered during the process of inspection for quality, which are (1) purity, (2) foreign matter. (3) defectives and (4) moisture content.A fuzzy weight value for each criterion was used to build the fuzzy hierarchical structure for the total evaluation of milled rice.

Table II shows the weights and details of each criterion. In this case study, 20 sample alternatives were used for the weights against each criterion. The data sample and total evaluation are as tabulated in Table II. The values for each criterion were assigned in a straightforward manner based on an intensity of compliance scale, stated in Table I.

Table IIItabulates the findings the weights obtained from regression model (4). where $a_i$ and $h_i$ denote a weight and its width of attribute $c_i$. The evaluations $c_1$ to $c_4$ in Table IIare the ones of attribute obtained from the experts. From the results exemplified in Table III, we conclude that in the expert judgment. (1) purity. (2) foreign matter. and (3) defectives attributes are evaluated important ones. followed by (0.93,0.000). (0.000, 0.224),and (0.07, 0.06). respectively. Other attribute are not strongly considered. This indicates that the attribute is also important and provide the flexibility covering from 0 to 0.224. Therefore, the result indicates that experts should place stress also for decision of foreign matter judgment. This weight value yielded from the fuzzy regression model is helpful to assist the grading process with minimal monitoring by human experts.

TABLE II
DATA SAMPLESAND TOTAL EVALUATION

| Sample | $(y_j, d_j)$ | $c_1$ | $c_2$ | $c_3$ | $c_4$ |
|---|---|---|---|---|---|
| A1 | (9,0.2) | 9 | 5 | 9 | 5 |
| A2 | (9,0.1) | 9 | 5 | 8 | 6 |
| A3 | (8,0.2) | 8 | 8 | 5 | 4 |
| A4 | (5,0.1) | 3 | 8 | 4 | 4 |
| A5 | (6,0.1) | 5 | 8 | 4 | 6 |
| A6 | (7,0.2) | 6 | 5 | 8 | 3 |
| A7 | (8,0.2) | 7 | 7 | 2 | 3 |
| A8 | (8,0.1) | 7 | 6 | 3 | 3 |
| A9 | (6,0.1) | 5 | 7 | 3 | 5 |
| A10 | (5,0.1) | 5 | 5 | 7 | 6 |
| A11 | (8,0.2) | 7 | 5 | 7 | 5 |
| A12 | (7,0.1) | 6 | 5 | 3 | 3 |
| A13 | (5,0.1) | 4 | 8 | 3 | 6 |
| A14 | (5,0.1) | 4 | 5 | 7 | 8 |
| A15 | (6,0.1) | 5 | 3 | 6 | 4 |
| A16 | (7,0.2) | 6 | 4 | 7 | 5 |
| A17 | (4,0.1) | 3 | 3 | 8 | 6 |
| A18 | (5,0.1) | 4 | 4 | 4 | 3 |
| A19 | (6,0.1) | 5 | 3 | 7 | 5 |
| A20 | (8,0.1) | 7 | 5 | 8 | 2 |

TABLE III
WEIGHTS OF CRITERIA

| weight | width |
|---|---|
| $a_1 = 0.93$ | $h_1 = 0.00$ |
| $a_2 = 0.00$ | $h_2 = 0.22$ |
| $a_3 = 0.07$ | $h_3 = 0.06$ |
| $a_4 = 0.000$ | $h_4 = 0.07$ |

where each weight $c_i = (a_i, h_i)$ for $i = 1,2,...,4$.

## VII. CONCLUSIONS

Information extraction is importance task in decision making. However such information is not easily obtained. or requires additional method. Meanwhile, human expertise is usually involved in decision making. The judgment experience and knowledge of these experts is unique to each person. However, better understanding of this judgment knowledge, which can be represented by weights of attribute during a decision making process. can be useful for facilitating the decision making process with minimal evaluation input from human experts. This uncertainty element is important, as the judgment evaluations strongly involve individual human preferences. Our work described in this paper reveals that fuzzy regression is useful to and can be effectively used to better facilitate the weight extraction process during the product quality inspection task.

## ACKNOWLEDGMENT

he author expresses her appreciation to the University un Hussein Onn Malaysia (UTHM) and Research .cculturation Grant Scheme (RAGS) Vot R044.

## REFERENCES

[1] J. Watada and W. Pedrycz, "A Fuzzy Regression Approach to Acquisition of Linguistic Rules," In: W. Pedrycz, A. Skowron, V. Kreinovich (eds), John Wiley & Sons Ltd, Chapter 32, pp. 719-730. 2008

[2] T.L. Saaty, *Multicriteria decision making: The analytic hierarchy process*, RWS Publications, Pittsburgh, PA, 1990.

[3] Nureize and J.Watada (2010) A fuzzy regression approach to a hierarchical evaluation model for oil palm fruit grading, Fuzzy Optimization and Decision Making, 9(1), pp. 105-122

[4] Chang, Y. H., & M Ayyub, B. (2001). Fuzzy regression methods–a comparative assessment. Fuzzy sets and systems, 119(2), 187-203.

[5] Fahmi, Z., Samah, B. A., & Abdullah, H. (2013). Paddy Industry and Paddy Farmers Well-being: A Success Recipe for Agriculture Industry in Malaysia. Asian Social Science, 9(3), p177.

[6] Daño, E. C., & Samonte, E. D. (2005). Public sector intervention in the rice industry in Malaysia. State intervention in the rice sector in selected countries: Implications for the Philippines.

[7] Höppner, F., & Klawonn, F. (2003). Improved fuzzy partitions for fuzzy regression models. International journal of approximate reasoning, 32(2), 85-102.

[8] Yang, M. S., & Ko, C. H. (1997). On cluster-wise fuzzy regression analysis. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 27(1), 1-13.

[9] Shapiro, A. F. (2005). Fuzzy regression models. Article of Penn State University.

[10] Song, K. B., Baek, Y. S., Hong, D. H., & Jang, G. (2005). Short-term load forecasting for the holidays using fuzzy linear regression method. Power Systems, IEEE Transactions on, 20(1), 96-101.

[11] Pasha, E., Razzaghnia, T., Allahviranloo, T., Yari, G., & Mostafaei, H. R. (2007). Fuzzy linear regression models with fuzzy entropy. Applied Mathematical Sciences, 1(35), 1715-1724.

[12] Abdollahzadeh, S., Rezaei, M., Dabbaghian, M., & Norouzi, A. (2010, April). A goal programming model for computation of fuzzy linear regression with least error.In Computer Engineering and Technology (ICCET), 2010 2nd International Conference on (Vol. 7, pp. V7-493). IEEE.

[13] J. Watada, "Trend of fuzzy multivariant analysis in management engineering," In:R. Khosla et al. (Eds.): *KES 2005, LNAI 3682*, Springer-Verlag Berlin Heidelberg, pp. 1283–1290, 2005.

[14] T.L. Saaty, *The analytic hierarchy process*, McGraw-Hill, New York, NY, 1980.

[15] H. Tanaka, and J. Watada, "Possibilistic linear systems and their Application to the linear regression model," *Fuzzy Sets and Systems*, vol.27, no.3, pp.275-289, 1988.

[16] S.V., Huffel, andJ. Vandewalle, "Algebraic relationships between classical regression and total least-squares estimation,"*Linear Algebra and its Applications*, vol. 93, 149-160, 1987.

[17] D. C., Montgomery, E. A., Peck, & , G. G., Vining Introduction to linear regression analysis, 821, Wiley, 2012.

[18] K.Dalkir, Knowledge Management in Theory and Practice, Elsevier Science, 2005.

[19] G. Khoshsima, C. Lucas, and A. Mohaghar, Assessing Knowledge Management with Fuzzy Logic, Practical Aspects of Knowledge Management, *Lecture Notes in Computer Science* 3336, 2004, pp 425-432

[20] J. Zhai, Y. Liang, J. Jiang, and Y. Yu, Fuzzy Ontology Models Based on Fuzzy Linguistic Variable for Knowledge Management and Information Retrieval Intelligent Information Processing IVIFIP – The International Federation for Information Processing, 288, 2008, pp 58-67

[21] G.Kaur and B. Verma. Measurement standards based grading of rice kernels by separating touching kernels for embedded imaging applications, *International Journal of Electronics, Communication & Instrumentation Engineering Research and Development*, 3 (1),2013, 127-134

[22] J. S., Aulakh, and V.K. Banga, Grading of rice grains by image processing, *International Journal Of Engineering Research & Technology*, 1(4), 2012.

[23] C.S. Silva, and U. Sonnadara, Classification of Rice Grains Using Neural Networks, *Proceedings of Technical Sessions*, 2013.

[24] A. Nureize, Pest Activity Prognosis in Rice Fields Using Fuzzy Expert System Approach, Masters thesis, Universiti Utara Malaysia, 2004.