EURASIAN JOURNAL OF ACADEMIC RESEARCH
Innovative Academy Research Support Center
UIF = 8.1 | SJIF = 5.685 www.in-academy.uz

BUILDING A FUZZY REGRESSION HYBRID MODEL

Egamberdiev Nodir Abdunazarovich
Associate Professor of Tashkent University of Information Technology named after Muhammad al-Khwarizmi
Berdiev Maruf Ramshiddin ogli
Third year student of Tashkent University of Information Technology named after Muhammad al-Khwarizmi
Xafizadinov Usnatdin Nuratdin Uli
First year student of Tashkent University of Information Technology named after Muhammad al-Khwarizmi
https://www.doi.org/10.5281/zenodo.8262850

ARTICLE INFO

Received: 09th August 2023
Accepted: 17th August 2023
Online: 18th August 2023

KEY WORDS

Regression, Fuzzy Logic, Fuzzy Regression, Hybrid Model.

ABSTRACT

The integration of fuzzy logic and regression in the novel Fuzzy Regression Hybrid Model is presented in this article. This innovative framework addresses uncertainty and variability in data, enhancing predictive accuracy. Through empirical case studies, its efficacy is demonstrated across diverse domains, showcasing its potential for precise modeling and decision-making in complex and dynamic environments. The Fuzzy Regression Hybrid Model emerges as a valuable tool for tackling real-world challenges.

Data gap filling methods are based on heuristic procedures that are very complex and iterative in nature, which, in addition to the difficulty of introducing the quality of the recovery of missing elements, justify the reason for criticism. The following shortcomings are mainly criticized:

– calculating the parameters of the gap-filling algorithm based on the available data, which establishes a relationship between the observations;
– the deterioration of the quality of grades with the increase in the percentage of vacancies;
– inability to strictly study the properties of the algorithm;
– violation of the nature of data and the nature of output.

While we acknowledge that this criticism is valid, the problem of filling in the data gaps is a very complex one, and theoretically it is very unlikely that we will have a correct and uniform solution that has no subjective elements. less Therefore, any new approach to solving this problem differs from known methods in terms of new heuristics and raises doubts about non-contradiction, adequacy, and correctness of application [1-2].

A fundamentally different method from the listed methods is the processing of blank data, which is not directly related to filling blanks.

Indeed, when constructing statistical associations based on omitted-element sampling, more accurate parameter estimates given the observation gaps are more important than correct filling-in-the-blanks. Below is a solution to the problem based on the methods of constructing intermediate models presented in the previous chapter [5-8].

We introduce some additional definitions. K we $(1,2,...,n)$ specify the number of observations. We divide the entire sample in the hands of the researcher into (X, y) two parts (X_k, y_k), $k \in K_1$ general (X_l, y_l) with observations (X_1, y_1) and non-generalized $l \in K_2$ with observations (X_2, y_2), where K_1 - the number of numbers of general observations, K_2 - a set of numbers of observations consisting of space, $K = K_1 \bigcap K_2$, $K_1 \bigcap K_2 = \emptyset$; X_k, as before, k is the -number X of the -matrix k. We assume only that the matrix consists of missing elements. X

has a way to fill in the gaps in the data . h This means that X the -row and -position i of the matrix k can contain its h complement values . We find $x_{ki}^1, x_{ki}^2, ..., x_{ki}^h$ the minimum x_{ki}^- and maximum x_{ki}^+ value from them. Thus, each space in the matrix X is replaced by an interval, and this matrix $[X^-, X^+]$ is transformed into an interval matrix. That is, non-whitespace elements on initialization will have upper and lower bounds that overlap across the spans. Replacing spaces with spaces does not have to be done in the above way. This can be done on the basis of a meaningful analysis of the data involving the knowledge of experts, or on the basis of "soft" procedures that ensure that the value falls into an insufficiently wide range . Filling in the blanks in this way is based on the idea of giving this part of the sample a corrective, auxiliary role in relation to complex information.

of the independent variables with space X is replaced by the interval matrix, the problem of estimating the linear regression parameters is reduced to constructing a set of estimates.

this case $y^- = y^+ = y$, in particular, the following conditions must be met:

$$X_1^- a^1 - X_1^+ a^2 \le y$$

$$X_1^+ a^1 - X_1^- a^2 \ge y .$$

$X_1^- = X_1^+ = X_1$ taking into account that, then $X_1 a = y$ equality should be fulfilled. But if all $(x_{k1}, x_{k2}, ..., x_{km}, y_k)$, $k \in K_1$ vectors are linear arbitrary and $\dim K > m$, then such $\exists k \in K$ satisfies the following relation:

$$\sum_{i=1}^{m} a_i x_{ki} \ne y_k .$$

We will consider in detail all the cases that arise when solving the problem of estimating the parameters of a linear equation based on a sample replaced by an interval of spaces.

we assume that rank . $X_1 \ge m$ Given this assumption, it is possible to estimate the full-sample linear regression parameters. Come, we denote the MNM estimator defined $\hat{a}$ in the set $([X_2^-, X_2^+], y_2)$ corresponding to the intermediate sampling (X_1, y_1) by , the signs of the

coefficients of the equation $\hat{\alpha}_i, i=\overline{1,m}$ overlap with the signs of known and . We determine the value of the loss function for evaluation $J_1(\hat{a})$ by a, i.e

$$J_1(\hat{a}) = \sum_{k\in K_1} \left| y_k - \sum_{i-1}^{m} \hat{a}_i x_{ki} \right|.$$

a) quasi-evaluation is a solution to the following CHD problem:

$$X_k a + r_k - s_k = y_k,\ k \in K_1$$

$$\bar{X}_{2l}^{-} a - u_3^l \le y_l$$

$$\bar{X}_{2l}^{+} a + u_3^l \ge y_l,\ l \in K_2$$

$$a_i \ge 0,\ агар\ \hat{a} \ge 0\ бўлса$$

$$a_i \le 0,\ агар\ \hat{a} \le 0, i=\overline{1,m}\ бўлса$$

$$r \ge 0,\ s \ge 0,\ u_3 \ge 0,\ v_3 \ge 0$$

$$\sum_{k\in K_1} (r_k + s_k) + p \sum_{k\in K_1} (u_3^l + v_3^l) \to \min,\ 0 < p \le 1$$

,

b) This situation (X_1, y_1) indicates that the interval found in the sample is not suitable for MNM-estimation, which is of particular interest when there are gaps in sufficiently wide intervals. (X_2, y_2) To assess the level of such disparity, the following question should be solved:

$$\bar{X}_l^{-} a \le y_l$$

$$\bar{X}_l^{+} a \ge y_l,\ l \in K_2$$

$$a + u - v = \hat{a}$$

$$a_i \ge 0,\ агар\ \hat{a} \ge 0\ бўлса$$

$$a_i \le 0,\ агар\ \hat{a} \le 0, i=\overline{1,m}\ бўлса$$

$$u \ge 0,\ v \ge 0$$

$$J(u,v) = \sum_{i=1}^{m} (u_i + v_i) \to \min$$

In order to learn the features of object modeling based on data analysis methods, it is advisable to perform the following tasks. It can be a deeper understanding of the content of the considered process, studying the laws of interrelationships of factors that are important according to the content, trying to find similarities between a number of objects. In many cases, a mathematical model is developed in order to evaluate the effectiveness of one or another management influence, to make predictive calculations of the object's development based on it. If the object or process that is subject to research has a certain history, then this history can be explored using substantive and formal tools in building a predictive model. If the process is inertial in one way or another, then the general thesis that the rules of its previous development will be manifested in the future serves as the basis. The quality of

retrospective information processing is of particular importance. It should be used completely (since a series of observations often have a short length) and correctly in order not to make false conclusions about one or another aspect of the behavior of the object being studied.

Data analysis is an effective tool for processing various types of information in order to build a mathematical model of objects of various nature (the term "applied statistics" is often used, but it is not quite correct). In its framework, many methods of data processing designed to build mathematical models have been developed [3].

Data analysis is a universal tool for statistical modeling of various objects. The widespread use of this approach is based on a number of reasons. First, it determines the existence factor and form of the connection between the important indicators, taking into account the stochasticity of the connections. Secondly, the use of models creates conditions for in-depth analysis of the studied systems, in addition to the main options of prediction, it allows to develop many auxiliary options based on various hypotheses regarding the future development of modeling objects. Thirdly, these models are very effective in monitoring the practical application of management solutions as a result of the possibility of reflecting structural and dynamic changes. Fourthly, this approach makes it possible to introduce a range of updates and additions to the models without difficulty, without significant changes, thereby increasing the accuracy of the acceleration of the studied processes. Fifth, statistical models imply the use of an information base based on an existing statistical report in the form of a dynamic observation series of indicators during construction. In addition, the development and use of such models is based on the use of uncomplicated formal methods of modern applied statistics, the degree of processing of their theoretical, algorithmic and software aspects is very high. Finally, as will be shown later, mathematical models based on both statistical and specific experimental information can be constructed and applied.

If we deal with the uncertainty of observing objects for which this or that distribution law does not have a sufficient number of observations to correctly confirm, or which cannot be strictly called homogeneous, then the classical statistical choice is doesn't.

At the same time, we can judge that there is a demonstration of a certain law behind them, even without having a sufficient number of observations. We cannot estimate the parameters of this law with sufficient precision, but we can come to a certain agreement about the appearance of this law and the range of dispersion of the key parameters included in its mathematical definition.

One of the main problems of statistical modeling is the construction of models that allow determining causal relationships between variables in ambiguous a priori information. Such models are designed to quantify the effects of voluntary versus nonvoluntary variables. The solution to this problem is quite difficult. The inconsistency of the modeling results with the real relationships is due to a number of reasons, in particular, due to the violation of the main directions in the regression analysis. Often they are violated due to multicollinearity. Such a situation leads to the estimation of parameters with a large variance, which in many cases does not allow meaningful interpretation of them, for example, due to incorrect signs. It can be seen that a priori information is a powerful tool for constructing the type of regression models mentioned above. In particular, it reduces the effect of multicollinearity on parameter

estimates. But forming a priori information is not always an easy task due to its uncertainty. One of the existing approaches to solving this problem is to use fuzzy math methods.

Given a fuzzy linear regression model:

$$y = a_0 + a_1 x_1 + a_2 x_2 + ... + a_n x_n .$$

(1)

1. Here, let a_i be fuzzy numbers in triangular form (2.1):

$$\mu_A(a_i) = \begin{cases} \dfrac{x - \beta_i}{c_i} & x > \beta_i, \\ \dfrac{\beta_i - x}{c_i} & x \le \beta_i. \end{cases}$$

a_i A fuzzy number can be written in parameter form as:

$$a_i = (\beta_i, c_i) .$$

Here β_i is the center of the interval, s_i - interval width $c_i \ge 0$.

In that case , u - the parameters of the fuzzy number are determined as follows:

interval center:

$$a_y = \sum_j \beta_j x_{ij} = \beta^T x ,$$

interval width:

$$c_y = \sum_j c_j |x_{ij}| = c^T |x| .$$

For a fuzzy logic model to be correct, *it* must belong to an interval uncertainty whose true value of the fuzzy number is bounded as follows:

$$\begin{cases} \beta^T x - c^T |x| \le y, \\ \beta^T x + c^T |x| \ge y. \end{cases}$$

(2)

The solution to problem (2) can be expressed as follows:

of fuzzy coefficients such that (β_i, c_i) :

a) the examined y_k let the values belong to the evaluation interval;

b) Let the evaluation interval "sum of widths" be minimal.

These requirements can be reduced to the following linear programming problem (1.2):

$$\begin{cases} \sum_k \sum_j c_j |x_{kj}| \to \min, \\ \sum_i \beta_i x_{ki} - \sum_i c_i |x_{ki}| \le y_k, \\ \sum_i \beta_i x_{ki} + \sum_i c_i |x_{ki}| \ge y_k. \end{cases}$$

EURASIAN JOURNAL OF ACADEMIC RESEARCH
Innovative Academy Research Support Center
UIF = 8.1 | SJIF = 5.685 www.in-academy.uz

2. (2.1) in the fuzzy linear regression model a_i let the coefficients be given by Gaussian fitness functions:

$$\mu_G(x) = e^{-\frac{1}{2}\frac{(x-\beta)^2}{c^2}}.$$

Such a fuzzy number is given using parameters of the following form:

$$\gamma = (\beta, c),$$

here βis the center of the fuzzy number, s is αthe width of the interval in the section of the fuzzy number.

The matter is put as follows:

(β_i, c_i) find coefficients a_i with parameters such that

a) y_k *let* the value belong to its evaluation set αwith a degree not less than;

b) αlet the evaluation interval be minimal.

αlevel rating interval:

$$d_\alpha = y_2 - y_1 = 2(y_2 - \beta).$$

Here it $y_2 - \beta$ is found from the following condition:

$$\alpha = \exp\left\{-\frac{1}{2}\frac{(y_2-\beta)^2}{c^2}\right\} \Rightarrow d_\alpha = 2(y_2-\beta) = 2c\sqrt{-2\ln\alpha}.$$

for the *k -th sample:*

$$d_\alpha^k = 2c_k\sqrt{-2\ln\alpha}.$$

Then we can express the objective function in the following form:

$$\min\sum_{k=1}^{m} d_\alpha^k = \min\sum_{k=1}^{m} 2c_k\sqrt{-2\ln\alpha} = 2\sqrt{-2\ln\alpha}\sum_{k=1}^{m}\sum_{i=1}^{n} c_i|x_{ik}|.$$

According to condition 1:

$$\mu(y_k) \ge \alpha \Rightarrow \exp\left\{-\frac{1}{2}\frac{(y_k-\beta_k)^2}{c_k^2}\right\} \ge \alpha \Rightarrow \begin{cases} \beta_k + c_k\sqrt{-2\ln\alpha} \ge y_k, \\ \beta_k - c_k\sqrt{-2\ln\alpha} \le y_k. \end{cases}$$

$$\beta_k = \sum_{i=1}^{n}\beta_i x_{ki};\ c_k = \sum_{i=1}^{n} c_i|x_{ki}|,$$

taking into account that, we solve the following linear programming problem to find the coefficients of the fuzzy model:

EURASIAN JOURNAL OF ACADEMIC RESEARCH
Innovative Academy Research Support Center
UIF = 8.1 | SJIF = 5.685 www.in-academy.uz

$$\begin{cases} \sum_{k=1}^{m}\sum_{j=1}^{n} c_j \left|x_{kj}\right| \to \min, \\ \sum_{i=1}^{n} \beta_i x_{ki} + \sqrt{-2\ln\alpha}\sum_{i=1}^{n} c_i \left|x_{ki}\right| \geq y_k, \\ \sum_{i=1}^{n} \beta_i x_{ki} - \sqrt{-2\ln\alpha}\sum_{i=1}^{n} c_i \left|x_{ki}\right| \leq y_k. \end{cases}$$

3. Let the relevance function be given in the following form:

$$\mu_Q(x) = \frac{1}{1+\left(\frac{x-\beta}{c}\right)^2}.$$

Such fuzzy numbers are given in the form of the following parameters:

$Q = (\beta, c)$,

here β- the center of the unclear number; *s* is the width of the function.

We look for the solution of the problem in the following way: (β_i, c_i) parameter such that a_i find the coefficients so that

a) y_k *let* the value belong to its evaluation set αwith a degree not less than ;

b) αthe evaluation interval of the level should be minimal.

αlevel rating interval:

$$d_\alpha = 2(y_2 - \beta).$$

Here it $y_2 - \beta$ is found from the following condition:

$$\alpha = \frac{1}{1+\left(\frac{y_2-\beta}{c}\right)^2} \Rightarrow \quad \alpha = \frac{c^2}{c^2+(y_2-\beta)^2} \Rightarrow$$

$$c^2 + (y_2-\beta)^2 = \frac{c^2}{\alpha} \Rightarrow (y_2-\beta)^2 = \frac{c^2(1-\alpha)}{\alpha} \Rightarrow$$

$$(y_2-\beta) = c\sqrt{\frac{(1-\alpha)}{\alpha}}.$$

Then the objective function can be expressed as:

$$\min\sum_{k=1}^{m} d_\alpha^k = \min\sum_{k=1}^{m} 2c_k\sqrt{\frac{(1-\alpha)}{\alpha}}; \quad c_k > 0 \Rightarrow \sum_{k=1}^{m}\sum_{i=1}^{n} c_i \left|x_{ik}\right| \to \min.$$

According to condition 1:

$$\mu(y_k) \geq \alpha \Rightarrow \frac{1}{1+\left(\frac{y_k-\beta_k}{c}\right)^2} \geq \alpha \Rightarrow \begin{cases} \beta_k + c_k\sqrt{\frac{(1-\alpha)}{\alpha}} \geq y_k, \\ \beta_k - c_k\sqrt{\frac{(1-\alpha)}{\alpha}} \leq y_k. \end{cases}$$

$$\beta_k = \sum_{i=1}^{n} \beta_i x_{ki};\ c_k = \sum_{i=1}^{n} c_i |x_{ki}|$$

taking into account that, we reduce the problem of finding fuzzy model coefficients to solving the following linear programming problem. This problem of linear programming is solved by simplex method or genetic algorithm [4-5].

$$\begin{cases} \sum_{k=1}^{m}\sum_{j=1}^{n} c_j |x_{kj}| \to \min, \\ \sum_{i=1}^{n} \beta_i x_{ki} + \sqrt{\frac{(1-\alpha)}{\alpha}} \sum_{i=1}^{n} c_i |x_{ki}| \geq y_k, \\ \sum_{i=1}^{n} \beta_i x_{ki} - \sqrt{\frac{(1-\alpha)}{\alpha}} \sum_{i=1}^{n} c_i |x_{ki}| \leq y_k. \end{cases}$$

Estimating the parameters of the regression equations included in a certain statistical model involves solving the problem of structural or joint estimation of the parameters of the system of equations. In this case, if the system is not recursive, the overall system parameter estimates will differ from the estimates calculated separately for each regression. An intermediate stage of structural assessment is the acquisition of additional a priori information. Statistical models, along with linear equations, include nonlinear regression equations and have a large size, which causes certain difficulties in solving the problem of specifying parameters of models, like structural estimation methods. One approach to solving this problem is the use of multilayer neural networks.

In conclusion, the Fuzzy Regression Hybrid Model exemplifies the synergy between fuzzy logic and regression, offering a robust solution for intricate data scenarios. This study highlights its capacity to mitigate uncertainty, adapt to dynamic trends, and provide accurate predictions. The model's success in various empirical cases underscores its potential to revolutionize decision-making and modeling across diverse industries. As a versatile tool, the Fuzzy Regression Hybrid Model proves invaluable for addressing complex real-world challenges and propels the field towards more accurate and informed analyses.

References:

1. P. Havali, J. Banu. Deep Convolutional Neural Network for Image Classification on CUDA Platform, ScienceDirect, 2019, Pages 99-122.
2. Dilnoz Muhamediyeva, Nadir Egamberdiyev. An application of Gauss neutrosophic numbers in medical diagnosis // International Conference on Information Science and Communications Technologies: Applications, Trends and Opportunities http://www.icisct2021.org/ ICISCT 2021, November 3-5, 2021.

3. Kamilov M.M., Khujaev OK, Egamberdiev NA The method of applying the algorithm of calculating grades for finding similar diagnostics in medical information systems, International Journal of Innovative Technology and Exploring Engineering, 8-6S, pp. 722-724.
4. Muhamediyeva DT, Jurayev Z.Sh., Egamberdiyev NA, Qualitative analysis of mathematical models based on Z-number // Proceedings of the Joint International Conference STEMM: Science – Technology – Education – Mathematics – Medicine. May 16-17, 2019, Tashkent, pp. 42-43.
5. Egamberdiyev NA, FUZZY REGRESSION ALGORITHM FOR CLASSIFICATION OF WEAKLY FORMED PROCESSES // SCIENCE AND PRACTICE: IMPLEMENTATION TO MODERN SOCIETY MANCHESTER, GREAT BRITAIN, 26-28.12.2020.
6. D.Mukhamediyeva, N.Egamberdiev, ALGORITHM OF CLASSIFICATION OF MEDICAL OBJECTS ON THE BASIS OF NEUTROSOPHIC NUMBERS, Proceedings of the 4th International Scientific and Practical Conference SCIENCE, EDUCATION, INNOVATION: TOPICAL ISSUES AND MODERN ASPECTS TALLINN, ESTONIA, 4- 5.10.2021, pp. 374-380.
7. Mukhamedieva DT, Egamberdiev NA, Zokirov J.Sh., Mathematical support for solving the classification problem using neural network algorithms // Turkish Journal of Computer and Mathematics Education. Vol.12 No.10 (2021).
8. DTMukhamedieva and NAEgamberdiev, APPROACHES TO SOLVING OPTIMIZATION TASKS BASED ON NATURAL CALCULATION ALGORITHMS, Scientific-technical journal, 3(2) 2020, pp. 58-67.
9. NAEgamberdiev, OTXolmuminov, KhROchilov, ANALYSIS OF CLASSICAL MODELS OF CLASSIFICATION OF SLOWLY FORMED PROCESSES, International Scientific-Online Conference: SOLUTION OF SOCIAL PROBLEMS IN MANAGEMENT AND ECONOMY", Spain, October 7, 2022, pp. 12-16.
10. NAEgamberdiev, OTXolmuminov, Khrochilov, CHOOSING AN EFFICIENT ALGORITHM FOR SOLVING THE CLASSIFICATION PROBLEM, International Scientific Online Conference: THEORETICAL ASPECTS IN THE FORMATION OF PEDAGOGICAL SCIENCES, October 10, 2022, pp.154-158.
11. D. Muhamedieva, N. Egamberdiev, O. Kholmuminov, APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR CREDIT RISK ASSESSMENT, "Science and innovation" international scientific journal. 2022, No. 6. Pages 388-395.
12. D. Muhamediyeva, N. Egamberdiyev, An application of Gauss neutrosophic numbers in medical diagnosis, International Conference on Information Science and Communications Technologies ICISCT 2021, Tashkent, Uzbekistan, 2021, pp. 1-4.
13. D. Muhamediyeva, N. Egamberdiyev, A. Bozorov, FORECASTING RISK OF NON-REDUCTION OF HARVEST, Proceedings of the 2nd International Scientific and Practical Conference, SCIENTIFIC COMMUNITY: INTERDISCIPLINARY RESEARCH, Hamburg, Germany, 26-28.01.2021, pp. 694-698.
14. F. Nuraliev, O. Narzulloev, N. Egamberdiev, S. Tastanova, V Mejdunarodnaya nauchno-prakticheskaya konferencija, RECENT SCIENTIFIC INVESTIGATION, Oslo, Norway, April 26-28, 2022, c. 447-451.

15. Mukhamedieva D.T., Egamberdiev N.A., Podkhody k resheniyu zadach optimizatsii na osnove algoritmov prirodnyx vychisleniy, Scientific and technical journal of Fergana Polytechnic Institute, 2020, Volume 24, No. 2, c. 75-84.
16. NAEgamberdiyev, MMKamilov, A.Sh. Hamroyev, Development of an algorithm for determining the system of dimensional fixed basis sets for educational selections, Muhammad al-Khorazmi Avlodali, 1(7) 2019, pp. 45-48.
17. Khojayev OQ, NAEgamberdiyev, Sh.N. Saidrasulov, Algorithm for choosing an effective method for solving the problem of classification, Information Communications: Networks, Technologies, Solutions. 1(49) 2019. Quarterly Scientific and Technical Journal, pp. 39-43.