

13. Galovic Ya. C++17 STL. Standartnaya biblioteka shablonov. Sankt-Peterburg: Piter, 2018. 432 p.
14. Analiz svoystv lineynoy sistemy metodom iskusstvennykh bazisnykh matric / Kudin V. I., Lyashko S. I., Hritonenko N. M., Yacenko Yu. P. // Kibernetika i sistemnyy analiz. 2007. Issue 4. P. 119–127.
15. Bohaienko V. O., Kudin V. I., Skopetskiy V. V. Analysis of computational schemes for basic matrix method // Komp'yuternaya matematika. 2009. Issue 2. P. 3–13.
16. Bogaenko V. A., Kudin V. I., Skopeckiy V. V. Analiz vychislitel'nykh skhem modelirovaniya processov geogidrodinamiki // Probl. upr. i informatiki. 2009. Issue 4. P. 62–72.
17. Bogaenko V. A., Kudin V. I., Skopeckiy V. V. Ob osobennostyakh organizatsii vychisleniy na osnove metoda bazisnykh matric // Kibernetika i sistemnyy analiz. 2012. Vol. 48, Issue 4. P. 146–155.
18. Bogainenko V., Kudin V. Building preconditioners using basis matrix method // International journal Information Content and Processing. 2014. Vol. 1, Issue 2. P. 182–187.
19. Knut D. Iskustvo programmirovaniya. Vol. 2. 3-e izd. Moscow: Izdatel'skiy dom «Vil'yams», 2000. 788 p.
20. Krendall R., Pomerans K. Prostye chisla: Kriptograficheskie i vychislitel'nye aspekty. Moscow: URSS, 2011. 664 p.
21. Straustrup B. Yazyk programmirovaniya C++. Special'noe izdanie. Sankt-Peterburg-Moscow: «Nevskiy dialekt» - «BINOM», 2006. 1104 p.
22. Kudin V. I., Onotskiy V. V. Rozvynennia tekhnolohiyi dovhoi aryfmetryky pry pobudovi alhorytmiv doslidzhennia zadachi liniynoho prohramuvannia // Zhurnal obchysliuvalnoi ta prykladnoi matematyky. 2011. Issue 1. P. 77–84.

Проведена модернізація методів ідентифікації стану об'єктів в умовах нечітких вхідних даних, описаних своїми функціями належності. Обраний напрямок вдосконалення традиційних методів пов'язаний із принциповими особливостями вирішення цього завдання в реальних умовах малої вибірки вхідних даних. У цих умовах для розв'язання задачі ідентифікації стану доцільно перейти до менш вибагливої в інформаційному відношенні технології опису вихідних даних, заснованої на математичному апараті нечіткої математики. Цей перехід зажадав розробки нових формальних методів вирішення конкретних завдань. При цьому для багатовимірного дискримінантного аналізу розроблено методіку розв'язання нечіткої системи лінійних алгебраїчних рівнянь. Для вирішення завдання кластеризації запропонована спеціальна процедура порівняння нечітких відстаней між об'єктами кластеризації і центрами групування. Обраний напрямок вдосконалення традиційного методу регресійного аналізу визначено неможливістю використання класичного методу найменших квадратів в умовах, коли всі змінні описані нечітко. Ця обставина привела до необхідності побудови спеціальної двохкрокової процедури вирішення завдання. При цьому реалізується мінімізація лінійної комбінації міри видалення шуканого рішення від модального і міри компактності функції приналежності пояснювальної змінної. Технологія нечіткого регресійного аналізу реалізована в важливому для практики випадку, коли вихідні нечіткі дані описані загальними функціями приналежності (L–R) типу. При цьому отримано аналітичний розв'язок задачі у вигляді розрахункових формул. В результаті обговорення показано, що модернізація класичних методів рішення задачі ідентифікації стану з урахуванням нечіткого характеру представлення вихідних даних дозволила проводити ідентифікацію об'єктів в реальних умовах малої вибірки нечітких вихідних даних

Ключові слова: нечіткі багатовимірний дискримінантний, кластерний, регресійний аналізи, технології зведення нечітких задач до чітких

UDC 519.85

DOI: 10.15587/1729-4061.2019.157085

IDENTIFICATION OF THE STATE OF AN OBJECT UNDER CONDITIONS OF FUZZY INPUT DATA

S. Semenov

Doctor of Technical Sciences, Professor*

O. Sira

Doctor of Technical Sciences, Professor**

S. Gavrylenko

PhD, Associate Professor*

N. Kuchuk

PhD

Department of Theoretical and Applied Systems Engineering

V. N. Karazin Kharkiv National University Svobody sq., 4, Kharkiv, Ukraine, 61022

E-mail: nina_kuchuk@ukr.net

*Department of Computer Science and Programming***

Department of Distributed Information Systems and Cloud Technologies*

***National Technical University «Kharkiv Polytechnic Institute»

Kyrpychova str., 2, Kharkiv, Ukraine, 61002

1. Introduction

Let us state the general principles for solving the problem of identification of the state of an object. Information base

is formed according to the results of measuring the values of a set of controllable parameters (features) of an object. Identification technologies provide linkage between these values and the state of an object. To solve this problem, a

number of special mathematical methods are traditionally used: multidimensional discriminant analysis, clustering, and regression analysis. However, the application of these techniques is significantly complicated, when the source data for identification of objects are determined fuzzily [1, 2]. The absence of the corresponding mathematical apparatus that makes it possible to solve the problem under these conditions, determines the relevance of the research.

2. Literature review and problem statement

The technology of multidimensional discriminant analysis is as follows [3]. Let an observation object be in one of two states H_1 and H_2 . The state of an object is determined by values p of indicators x_1, x_2, \dots, x_n . The numeric values of the controlled indicators are supposed to be normally distributed random magnitudes. In this case, mathematical expectations are determined by vector $M_1=(m_{11}, m_{12}, \dots, m_{1p})$, if an object is in state H_1 , and by $M_2=(m_{21}, m_{22}, \dots, m_{2p})$, if an object is in state H_2 . It is also assumed that the elements of the matrix of coefficients of correlation between indicators $K=(k_{ij}), i=1, 2, \dots, p$ and $j=1, 2, \dots, p$ do not depend on the state of an object. To assess the state of an object by results of measurements of the controlled indicators, the discriminant function

$$Z = \sum_{i=1}^p a_i x_i.$$

is used.

Unknown coefficients $a_i, i=1, 2, \dots, p$ of this function are derived as a result of solving the system of linear equations

$$\sum_{i=1}^p k_{1i} a_i = m_{11} - m_{21}, \dots, \sum_{i=1}^p k_{pi} a_i = m_{1p} - m_{2p}. \quad (1)$$

Using the results of the solution of system (1), the values are calculated:

$$\zeta_1 = \sum_{i=1}^p a_i m_{1i}, \quad \zeta_2 = \sum_{i=1}^p a_i m_{2i}, \quad C = \frac{1}{2}(\zeta_1 + \zeta_2).$$

Now, the decisive rule is stated: an object is in state \hat{H}_1 , if for a specific set of values of controlled indicators $\hat{x}_{11}, \hat{x}_{12}, \dots, \hat{x}_{1n}$ the corresponding value of discriminant function satisfies inequality:

$$\hat{Z} = \sum_{i=1}^p a_i \hat{x}_i < C,$$

otherwise – in state H_2 .

As shown in [4], selection of values of coefficients $a_i, i=1, 2, \dots, p$, values ζ_1, ζ_2 and C ensures a minimum of the total probability of confusing the states, equal to $p(H_1/H_2) + p(H_2/H_1)$.

We will note the general shortcomings of the traditional method of multidimensional discriminant analysis. First, only double-alternative diagnosis was implemented in the method. It is not enough when solving many practical problems [5]. Second, in the traditional method, discriminant surface is a hyperplane. The coefficients of its equation are found by the statistical characteristics of the two points that represent a subspace of the phase space of observations. In

this case, the error of diagnoses confusion can be very large [6]. The real accuracy of estimates of conditions using specific multifactor discriminant models is unpredictable and significantly depends on the nature and characteristics of sample data, their volume, uniformity, the sense of controlled indicators [7]. In addition, it should be noted that in a range of works, for example, in [8, 9], the assumption of Gaussian character of random observation values is used, which considerably limits their application areas. Another method for solving the problem of the objects set recognition is more reliable. Let us proceed to consideration of fuzzy clustering.

Let the results of measurements p of the indicators of each object make up a set of points of p -dimensional phase space. Cluster analysis technology makes it possible to split the source sets into m subsets (by the number of possible object types (H_1, H_2, \dots, H_m)). In this case, the points belonging to one subset – a cluster, in some selected (specified) sense are “close” to each other and “far” from the points of other subsets of clusters [10]. A lot of different methods of clustering are known. Most of them in different variants implement the following simple procedure that is described, for example, in [11]. The number of clusters is known a priori and the grouping center (that is, sets of coordinates of typical points for corresponding states of an object) is assigned for each of them. Now, we perform an iteration procedure, at each step of which the distances to the centers of cluster grouping are found for the successive distributed point and the shortest of these distances is selected for point joining. The most important element of the clustering technology is the procedure of comparison of distances. There are also other ways of implementing this procedure [12, 13]. In all cases, it is assumed that the coordinates of points and grouping centers are measured precisely (or the estimation error is distributed normally). This limits the range of application of these methods in the context of fuzzy source data.

The merit of the clustering method is the ease of implementation and unambiguous interpretation of the results. The drawback of this method is low informative value. In fact, the fact that a point belongs to a cluster does not contain any important information about the location of this point in the cluster, that is, whether this point is at the center of the cluster, or near the boundary with a neighboring cluster. Consider a more informative method for identification that is based on regression analysis.

Regression analysis is a powerful, effective method that describes the relationship between some of the selected indicators of a control object and its directly measured characteristics, indicators, and parameters. A general drawback of this method is the lack of the grounded choice of controlled indicators and the procedure of determining the coefficients of the model (1).

Consider the well-posed statement of the problem on regression analysis. Controlled indicators (x_1, x_2, \dots, x_p) , presumably affecting the resulting indicator of the quality of functioning of object y , are selected by any well-grounded method. The relationship between the explaining variables (x_1, x_2, \dots, x_p) and resulting variable y is described by Kolmogorov-Gabor polynomial, which in its simplest form, is:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_p x_p + \varepsilon. \quad (2)$$

To find the unknown coefficient of mode a_0, a_1, \dots, a_p a series of n experiments is carried out. In this case, every

experience $X_j = (x_{j1}, x_{j2}, \dots, x_{jp})$ is put in correspondence its result $y_j, j=1, 2, \dots, n$, that is:

$$y_j = a_0 + a_1x_{j1} + a_2x_{j2} + \dots + a_px_{jp} + \epsilon_j, \quad j=1, 2, \dots, n.$$

In a matrix form, this ratio is as follows $Y = HA + E$, where:

$$H = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}; \quad A = \begin{pmatrix} a_0 \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{pmatrix}; \quad Y = \begin{pmatrix} y_0 \\ \cdot \\ \cdot \\ \cdot \\ y_p \end{pmatrix}; \quad E = \begin{pmatrix} \epsilon_0 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_p \end{pmatrix}.$$

In the classic theory of regression analysis (a Gauss-Markov model), it is assumed that random measurement errors y_j in each experience are not correlated and normally distributed with zero mathematical expectation and known constant dispersion. In this case, estimates of unknown coefficients a_0, a_1, \dots, a_p are obtained by the least squares method, through minimizing criterion

$$J = (HA - Y)^T (HA - Y).$$

Vector A , minimizing this criterion is determined from ratio: $\hat{A} = (H^T A)^{-1} H^T Y$.

We will perform a brief analysis of the described methods for identification of states. The need for improvement of ideas, methods and technologies of identification, which has occurred in recent years, is linked to the formed understanding the inadequacy of the theoretical-probabilistic models of uncertainty for most actual problems of evaluation of the state of an object. The main causes that these models are unsatisfactory include a small source data sample, as well as the change of the conditions for functioning of a control object. Thus, in terms of a small sample of a priori source data [14], the hypothesis of normality of observed data cannot be either properly justified or rejected, which questions the legality of using the central limit theorem. In [15], for the same reason, the errors of statistical estimates of mathematical expectations and variances of the controlled indicators can be unpredictably large. This circumstance will inevitably lead to respectively large errors when solving the system of equations (1), estimates of coefficients of discriminant functions and, consequently, as a result, the identification error will be great. The approach that implies refusal from a priori assumptions about normality of the observed values of parameters of an object in favor of the model of fuzzy mathematics is natural in this situation [16]. This mathematical apparatus is much less sensitive to the sample volume and makes it possible to determine reliably the key structural elements of this theory – the membership function of numeric values of observed indicators even under conditions of a small source data sample. In this case, the simplest variant of the solution of the problem of diagnosing the state of an object using the technology of multidimensional discriminant analysis is calculation and use of theoretical-probabilistic analogues of statistical characteristics of observed magnitudes.

As regards a regression analysis, the transition to the description of source data in terms of fuzzy mathematics initiated the development of new technologies. In [17, 18], the membership function of the resulting variable, which is compared with experimental membership function, is

determined. The fundamental drawback of this approach is that the accuracy of estimation of the values of independent and dependent variables in practice differ considerably. That is why the result of solving this problem not necessarily will provide a minimum of total fuzziness of the described variable. This raises doubts as to the correctness of description of the relationship between explaining and explained variables. The absence of analysis of result proximity to the modal value of fuzzy explained variable, obtained by results of statistical treatment of source data in [19], decreases the effectiveness of the proposed method. Non-compact function of membership of a fuzzy value of the explained variable in [20] increased the error of the proposed solution.

3. The aim and objectives of the study

The aim of this study is to modernize traditional identification methods taking into consideration the fuzziness of source data.

To achieve this aim, it is necessary to solve the following problems:

- to develop a fuzzy method for discriminant analysis;
- to develop a method of fuzzy clustering;
- to develop an effective method for fuzzy regression analysis.

4. Modernization of methods for identification of the state of objects under conditions of fuzzy source data

Let us assume that according to the results of previous studies for each indicator of object x_i , a set of values for this indicator $(x_{i1}^{(1)}, x_{i2}^{(1)}, \dots, x_{i1}^{(1)})$ was determined for the case when an object is in state H_1 , and set $(x_{i1}^{(2)}, x_{i2}^{(2)}, \dots, x_{i1}^{(2)})$ for the case when an object is in state H_2 . According to these data, we will obtain the description, for example, of the triangle membership function of parameter x_i . In this case, we have:

$$\mu(x_i / H_1) = \begin{cases} 0, & x_i < b_{1i}, \\ \frac{x_i - b_{1i}}{m_{1i} - b_{1i}}, & b_{1i} < x_i < m_{1i}, \\ \frac{c_{1i} - x_i}{c_{1i} - m_{1i}}, & m_{1i} < x_i \leq c_{1i}, \\ 0, & x_i > c_{1i}; \end{cases}$$

$$\mu(x_i / H_2) = \begin{cases} 0, & x_i < b_{2i}, \\ \frac{x_i - b_{2i}}{m_{2i} - b_{2i}}, & b_{2i} < x_i < m_{2i}, \\ \frac{c_{2i} - x_i}{c_{2i} - m_{2i}}, & m_{2i} < x_i \leq c_{2i}, \\ 0, & x_i > c_{2i}, \end{cases}$$

where

$$b_{1i} = \min\{x_{i1}^{(1)}, x_{i2}^{(1)}, \dots, x_{i1}^{(1)}\}, \quad c_{1i} = \max\{x_{i1}^{(1)}, x_{i2}^{(1)}, \dots, x_{i1}^{(1)}\},$$

$$b_{2i} = \min\{x_{i1}^{(2)}, x_{i2}^{(2)}, \dots, x_{i1}^{(2)}\}, \quad c_{2i} = \max\{x_{i1}^{(2)}, x_{i2}^{(2)}, \dots, x_{i1}^{(2)}\},$$

$$m_{1i} = \frac{1}{l_1} \sum_{s=1}^{l_1} X_{is}^{(1)}; \quad m_{2i} = \frac{1}{l_2} \sum_{s=1}^{l_2} X_{is}^{(2)}.$$

We will calculate the values of the main theoretical-probabilistic characteristics of fuzzy magnitudes x_i , $i=1, 2, \dots, p$. Let us introduce functions:

$$\varphi_i = (x_i / H_1) = \frac{\mu(x_i / H_1)}{\int_{b_i}^{c_i} \mu(x_i / H_1) dx_i},$$

$$\varphi_i = (x_i / H_2) = \frac{\mu(x_i / H_2)}{\int_{b_i}^{c_i} \mu(x_i / H_2) dx_i}.$$

These functions are non-negative and the integral of them is equal to unity. That is why they can be interpreted as the density of probability of random magnitudes and used to calculate their mathematical expectations:

$$\hat{m}_{1i} = \int_{b_i}^{c_i} x_i \varphi_i(x_i / H_1) dx_i,$$

$$\hat{m}_{2i} = \int_{b_i}^{c_i} x_i \varphi_i(x_i / H_2) dx_i, \quad i=1, 2, \dots, p.$$

To calculate the estimates of the elements of the correlation matrix, we will determine:

$$\mu_a(x_i / H_1 \vee H_2) = \begin{cases} 0, & x_i \leq b_i, \\ \frac{x_i - b_i}{m_i - b_i}, & b_i < x_i \leq m_i, \\ \frac{c_{2i} - x_i}{c_{2i} - m_i}, & m_i < x_i \leq c_{2i}, \\ 0, & x_i > c_{2i}, \end{cases}$$

where

$$m_i = \frac{1}{l_1 + l_2} \left(\sum_{s=1}^{l_1} x_{is}^{(1)} + \sum_{s=1}^{l_2} x_{is}^{(2)} \right), \quad i=1, 2, \dots, p,$$

as well as:

$$\varphi_i = (x_i / H_1 \vee H_2) = \frac{\mu(x_i / H_1 \vee H_2)}{\int_{b_i}^{c_i} \mu(x_i / H_1 \vee H_2) dx_i}, \quad i=1, 2, \dots, p.$$

Then theoretic-probabilistic analogue of dispersion of indicator x_i for the whole set of its observations is equal to:

$$D[x_i] = \int_{b_i}^{c_i} x_i^2 \varphi(x_i / H_1 \vee H_2) dx_i.$$

Then

$$\hat{k}_{ri} = \frac{1}{(l_1 + l_2)(D[x_r]D[x_i])^{1/2}} \sum_{s=1}^{l_1+l_2} (x_{rs} - \mu_r)(x_{is} - \mu_i),$$

$$i=1, 2, \dots, p, \quad r=1, 2, \dots, p.$$

The derived estimates for mathematical expectation and correlation factors will be subsequently used for the standard scheme of calculation of set a_i , $i=1, 2, \dots, p$, by solving the system of linear equations (1) and subsequent actions during solution of the problem of diagnosing the object state.

Let us proceed to consideration of fuzzy clustering. Let the coordinates of the points (the results of measurement of the controlled indicators), as well as clusters grouping centers, be assigned in a fuzzy way by their membership functions. Then, the membership function of fuzzy distance for any pair (point – grouping center) can be obtained by the known rules for performing the operations over fuzzy numbers [21].

For example, membership function of the i -th coordinate of the k -th grouping center and membership function of the same coordinate of the j -th point are assigned by membership functions of the $(L-R)$ type:

$$\mu_{ik}(x_{ik}) = \begin{cases} L\left(\frac{\bar{x}_{ik} - x_{ik}}{\alpha_{ij}}\right), & x_{ik} \leq \bar{x}_{ik}, \\ R\left(\frac{x_{ik} - \bar{x}_{ik}}{\beta_{ik}}\right), & x_{ik} > \bar{x}_{ik}; \end{cases}$$

$$\mu_{ij}(x_{ij}) = \begin{cases} L\left(\frac{\bar{x}_{ij} - x_{ij}}{\alpha_{ij}}\right), & x_{ij} \leq \bar{x}_{ij}, \\ R\left(\frac{x_{ij} - \bar{x}_{ij}}{\beta_{ij}}\right), & x_{ij} > \bar{x}_{ij}, \end{cases}$$

where \bar{x}_{ik} is the modal value of the i -th coordinate of the k -th grouping center, \bar{x}_{ij} is the modal value of the i -th coordinate of the j -th point, α_{ik} , α_{ij} are the left fuzziness coefficients, β_{ik} , β_{ij} are the right fuzziness coefficients.

To calculate membership function of fuzzy distance between the k -th grouping center and the j -th point by the i -th coordinate, we will use the following rules of performing the operations over fuzzy numbers of the $(L-R)$ type [22]. Let

$$x_1 = \langle a_1, \alpha_1, \beta_1 \rangle, \quad x_2 = \langle a_2, \alpha_2, \beta_2 \rangle$$

be fuzzy number of the $(L-R)$ -type. Then parameters of fuzzy number

$$C = x_1 + x_2 = \langle a, \alpha, \beta \rangle$$

are found from formulas:

$$a = a_1 + a_2, \quad \alpha = \alpha_1 + \alpha_2, \quad \beta = \beta_1 + \beta_2;$$

parameters of fuzzy number

$$C = x_1 - x_2 = \langle a, \alpha, \beta \rangle$$

are found from formulas:

$$a = a_1 - a_2, \quad \alpha = \alpha_1 + \alpha_2, \quad \beta = \beta_1 + \beta_2;$$

parameters of fuzzy number

$$C = x_1 \cdot x_2 = \langle a, \alpha, \beta \rangle$$

are found from formulas:

$$a = a_1 \cdot a_2,$$

$$\alpha = |a_1| \alpha_2 + |a_2| \alpha_1 + \alpha_1 \alpha_2, \quad \beta = |a_1| \beta_2 + |a_2| \beta_1 + \beta_1 \beta_2;$$

parameters of fuzzy number $C=x_1/x_2$ (x_1, x_2 – positive numbers) are found from formulas:

$$a = a_1 / a_2, \quad \alpha = \frac{a_1\beta_2 + a_2\alpha_1}{a_2^2}, \quad \beta = \frac{a_1\alpha_2 + a_2\beta_1}{a_2^2}.$$

In this case, parameters of fuzzy distance by the i -th coordinate between the j -th point and the k -th grouping center and square of this distance are determined by ratios:

$$\begin{aligned} r_{ijk} &= \langle a_{ijk}^{(1)}, \alpha_{ijk}^{(1)}, \beta_{ijk}^{(1)} \rangle, \quad a_{ijk}^{(1)} = \bar{x}_{ik} - \bar{x}_{ij}, \\ \alpha_{ijk}^{(1)} &= \alpha_{ik} + \beta_{ij}, \quad \beta_{ijk}^{(1)} = \beta_{ik} + \alpha_{ij}; \\ (r_{ijk})^2 &= \langle a_{ijk}^{(2)}, \alpha_{ijk}^{(2)}, \beta_{ijk}^{(2)} \rangle, \quad \alpha_{ijk}^{(2)} = (\alpha_{ijk}^{(1)})^2, \\ \alpha_{ijk}^{(2)} &= a_{ijk}^{(1)}\alpha_{jki}^{(1)} + a_{ijk}^{(1)}\alpha_{jki}^{(1)} = 2(a_{ijk}^{(1)}\alpha_{jki}^{(1)}), \\ \beta_{ijk}^{(2)} &= a_{ijk}^{(1)}\beta_{jki}^{(1)} + a_{ijk}^{(1)}\beta_{jki}^{(1)} = 2(a_{ijk}^{(1)}\beta_{jki}^{(1)}). \end{aligned}$$

Now, parameters of membership function of fuzzy square of the distance between the k -th grouping center and the j -th point are equal to:

$$\begin{aligned} r_{jk}^2 &= \sum_{i=1}^p r_{ijk}^2 = \langle a_{jk}, \alpha_{jk}, \beta_{jk} \rangle, \\ a_{jk} &= \sum_{i=1}^p (a_{ijk}^{(1)})^2; \quad \alpha_{jk} = \sum_{i=1}^p \alpha_{ijk}^{(2)}; \\ \beta_{jk} &= \sum_{i=1}^p \beta_{ijk}^{(2)}. \end{aligned}$$

Further operations of clustering procedure are determined by the rule of comparison for pairs of fuzzy numbers. The strict approach to solving this problem is proposed in [22] and is implemented as follows. Let fuzzy numbers x and y be assigned by their membership functions $\mu(x)$ and $\mu(y)$. The degree of preference of number x to number y is determined from formula:

$$\eta(x, y) = \sup_{x > y} \min\{\mu(x), \mu(y)\},$$

And degree of preference of y to x is determined from formula:

$$\eta(y, x) = \sup_{x < y} \min\{\mu(y), \mu(x)\}.$$

Then number x is “larger” than y , if $\eta(x, y) > \eta(y, x)$, and number x is “smaller” than y otherwise. Practical implementation of this procedure is complicated. That is why different heuristic approaches are used in practice to solve the problem of comparison of fuzzy numbers [23, 24]. One of them is implemented as follows. Degrees of belonging of x and y to corresponding sets on the set of levels v_1, v_2, \dots, v_s are found for assigned membership functions $\mu(x)$ and $\mu(y)$. These values are determined through solving equations $\mu(x) = v_r, \mu(y) = v_r, r = 1, 2, \dots, s$. Based on results of solving these equations, we will obtain the set of pairs of their roots: $\{(x_{v_r}^{(1)}, x_{v_r}^{(2)})\}, \{(y_{v_r}^{(1)}, y_{v_r}^{(2)})\}$. Now we will consider that fuzzy number x is “larger” than y , of for all $v_r, r = 1, 2, \dots, s$, the inequality is satisfied: $x_{v_r}^{(1)} + x_{v_r}^{(2)} \geq y_{v_r}^{(1)} + y_{v_r}^{(2)}$, and at least one of them is strictly satisfied. If this condition is not satisfied, neither of numbers x and y has any advantage over the other.

We will note the shortcomings of the above approach. Firstly, it is not clear at how many levels it is necessary to perform v -section. Secondly, the approach can be difficult to implement. Thirdly, the described approach will have a specific result only in the case of an obvious advantage of one number over the other, for example, if there is no intersection of membership functions of compared numbers.

Owing to this, another more simple and reliable approach with the result, which is interpreted unambiguously, is proposed.

Let x and y be the fuzzy triangular numbers with membership functions:

$$\mu(x) = \begin{cases} 0, & x \leq b_x, \\ \frac{x - b_x}{m_x - b_x}, & b_x < x < m_x, \\ \frac{c_x - x}{c_x - m_x}, & m_x < x \leq c_x, \\ 0, & x > c_x; \end{cases}$$

$$\mu(y) = \begin{cases} 0, & y \leq b_y, \\ \frac{y - b_y}{m_y - b_y}, & b_y < y < m_y, \\ \frac{c_y - y}{c_y - m_y}, & m_y < y \leq c_y, \\ 0, & y > c_y. \end{cases}$$

We will find membership function of difference $z = x - y$:

$$\mu(z) = \begin{cases} 0, & z \leq b_z, \\ \frac{z - b_z}{m_z - b_z}, & b_z < z < m_z, \\ \frac{c_z - z}{c_z - m_z}, & m_z < z \leq c_z, \\ 0, & z > c_z, \end{cases}$$

where

$$b_z = b_x - b_y, \quad m_z = m_x - m_y, \quad c_z = c_x - c_y.$$

Now the original problem of comparing x and y is reduced to a simpler problem of comparison of fuzzy number z with zero.

We will introduce the rules of interpretation of the result of comparing fuzzy number z with zero:

- a) if $\min\{b_z, c_z\} > 0$, then $x > y$,
- b) if $\max\{b_z, c_z\} < 0$, then $x < y$,
- c) if

$$\min\{b_z, c_z\} < 0, \max\{b_z, c_z\} > 0, \text{ and } |\min\{b_z, c_z\}| > \max\{b_z, c_z\},$$

then $x < y$,

- d) if

$$\min\{b_z, c_z\} < 0, \max\{b_z, c_z\} > 0, \text{ and } |\min\{b_z, c_z\}| < \max\{b_z, c_z\},$$

then $x > y$.

Examples of possible results of subtraction operations are shown in Fig. 1–5. In all cases, graphic descrip-

tions of summands are shown in figures on the left, and the results of subtraction operation are shown on the right.

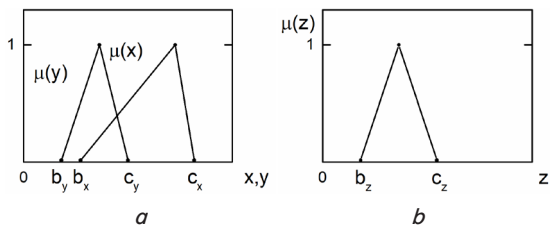


Fig. 1. Result of calculation of $x - y, x > y$ (rule a , variant 1):
 a – graphic descriptions of summands;
 b – result of calculation operation

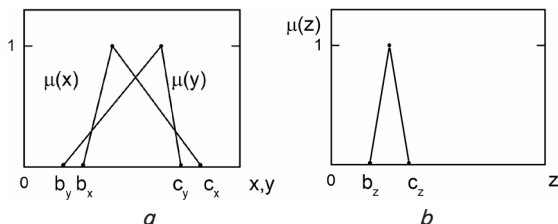


Fig. 2. Result of calculation $x - y, x > y$ (rule a , variant 2):
 a – graphic descriptions of summands; b – result of calculation operation

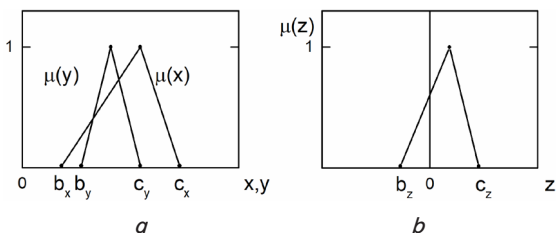


Fig. 3. Result of calculation $x - y, c_z > |b_z|, x > y$ (rule a):
 a – graphic descriptions of summands; b – result of calculation operation

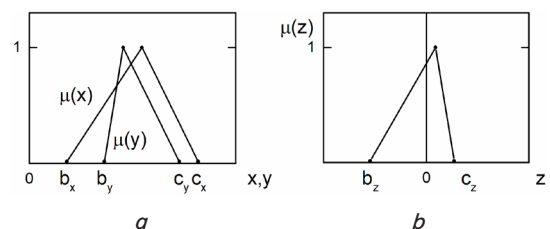


Fig. 4. Result of calculation $x - y, |b_z| > c_z, x < y$ (rule c):
 a – graphic description of summands; b – result of calculation operation

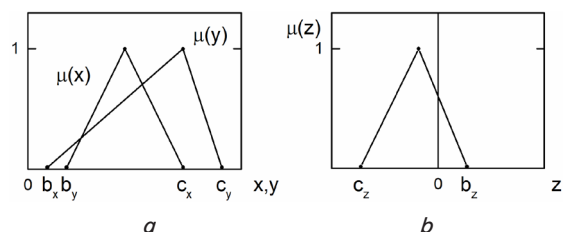


Fig. 5. Result of calculation $x - y, b_z < |c_z|, x < y$ (rule c):
 a – graphic description of summands; b – result of calculation operation

Thus, we obtain the following rules:

- 1) if the carrier of fuzzy result of subtracting is positive, then the minuend is larger than subtrahend;
- 2) if the carrier of the result is negative, then the minuend is smaller than the subtrahend;
- 3) if the carrier covers zero and its negative section is larger than the positive one, then the minuend is smaller than the subtrahend;
- 4) if the carrier covers zero and its negative section is smaller than the positive one, then minuend is larger than subtrahend.

Let us assume that the coordinates of grouping centers of objects are assigned. Then in the clustering problem, the shortest distance, determining the cluster, to which this point is necessary to join, is selected by the results of comparison of fuzzy distances from the next point to the cluster centers according to specified rules. Results of solving the clustering problem for a training set of objects are used to specify the coordinates of clusters grouping centers.

Let us proceed to solution of the third problem. The effective way of improving the quality of the solution of the problem of regression analysis, which ensures getting a solution satisfying two natural requirements, is proposed in [21]:

- 1) proximity of the result to the modal value of a fuzzy explained variable, obtained by the results of statistical treatment of the source data;
- 2) membership function of fuzzy value of the explained variable should be maximum compact.

Let us choose Gaussian membership function to describe fuzzy source data:

$$\mu(x_{ji}) = \exp \left\{ - \frac{(x_{ji} - x_{ji}^{(0)})^2}{2\delta_{ji}^2} \right\}, \quad j = 1, 2, \dots, n, \quad i = 1, 2, \dots, p.$$

The problem is solved in two stages [25]. At the first stage, the system of linear algebraic equations is composed relatively to the unknown values of coefficients $a_i, i = 1, 2, \dots, m$ of regression ratio (2). The coefficients of the composed system of linear algebraic equations (SLAE) are assigned by modal values $x_{ji}^{(0)}$ of explaining variables:

$$\begin{cases} a_0 + x_{11}^{(0)} a_1 + x_{12}^{(0)} a_2 + \dots + x_{1p}^{(0)} a_p = y_1, \\ a_0 + x_{21}^{(0)} a_1 + x_{22}^{(0)} a_2 + \dots + x_{2p}^{(0)} a_p = y_2, \\ \dots \\ a_0 + x_{n1}^{(0)} a_1 + x_{n2}^{(0)} a_2 + \dots + x_{np}^{(0)} a_p = y_n. \end{cases} \quad (3)$$

Solution to this system with the use of the least squares methods (LSM) (since system (3) is redetermined) gives set $a_i^{(0)}, i = 1, 2, \dots, p$ which assigns modal values of sought-for variables a [26].

Next, a set of fuzzy numbers is introduced

$$Z_j = \sum_{i=0}^p a_i x_{ji} - y_j, \quad j = 1, 2, \dots, n, \quad x_{j0}^{(0)} = 1,$$

fuzziness of which is determined by fuzziness of explaining variables x_{ji} . Membership function of fuzzy value z_j is determined from ratio:

$$\mu(Z_j) = \exp \left\{ - \left(z_j - \sum_{i=0}^m (a_i x_{ji}^{(0)}) \right)^2 / (2\delta_{ji}^2) \right\}.$$

Then compactness of the solution is determined by compactness of fuzzy numbers z_j , estimated by the area of the figure curve $\mu(z_j)$, which is equal to:

$$s(\mu(z_j)) = \int_{-\infty}^{\infty} \mu(z_j) dz_j.$$

At the second stage of problem solving, the compound criterion is stated [27]:

$$J(A) = \int_{-\infty}^{\infty} \mu(z_j) dz_j + \sum_{i=0}^m (a_i - a_i^{(0)})^2, \tag{4}$$

which is minimized by $a_i, i=1, 2, \dots, p$.

Thus, according to criterion (4), we estimate the set of coefficients of regression equation $a_i, i=1, 2, \dots, p$, which ensure minimum blurring (maximum compactness) of membership function of result and minimum deviation from modal set $a_i^{(0)}, i=1, 2, \dots, p$. A significant drawback of this approach is complexity of solving the system of equations, obtained in this case.

$$\frac{dJ(A)}{da_i} = 0, \quad i=1, 2, \dots, p,$$

which even in the simplest case, when fuzziness of source data is described in the Gaussian form, can be solved only numerically. Difficulties of solution increase additionally, if we use a general expression in the form of a function of the (L-R)-type to describe x_{ij} :

$$\mu(x_{ij}) = \begin{cases} ((x_{ij}^{(0)} - x_{ij})/\alpha_{ij})^2, & \text{if } x_{ij} \leq x_{ij}^{(0)}, \\ ((x_{ij}^{(0)} - x_{ij})/\beta_{ij})^2, & \text{if } x_{ij} \geq x_{ij}^{(0)}. \end{cases} \tag{5}$$

We will simplify expression for criterion (4), bearing in mind (5). Compactness of fuzzy number x_{ij} can be estimated by the sum of the left and right fuzziness coefficients $c_{ji} = \alpha_{ji} + \beta_{ji}, i=1, 2, \dots, p, j=1, 2, \dots, n$. The measure of compactness of number z_j will be equal to:

$$\eta_j(a_1, a_2, \dots, a_n) = \sum_{i=1}^p c_{ji}^2 a_i^2, \quad j=1, 2, \dots, n.$$

In this case, minimizing compound criterion (4) of the problem will take the form of:

$$\begin{aligned} J(A) &= \sum_{j=1}^n \sum_{i=1}^p c_{ji}^2 a_i^2 + \sum_{i=0}^m (a_i - a_i^{(0)})^2 = \\ &= \sum_{i=1}^p d_i a_i^2 + \sum_{i=0}^m (a_i - a_i^{(0)})^2 \Rightarrow \min. \end{aligned} \tag{6}$$

Minimum (6) will be found in the assumption that:

$$\sum_{i=0}^p a_i = a_0 + \sum_{i=1}^p a_i = 1, \quad a_0 = a_0^{(0)}. \tag{7}$$

Using the method undetermined Lagrange multipliers. We will introduce the Lagrangian function:

$$L(A) = \sum_{i=1}^p d_i a_i^2 + \sum_{i=0}^m (a_i - a_i^{(0)})^2 - \lambda \left(\sum_{i=1}^p a_i - 1 + a_0^{(0)} \right).$$

Then

$$\frac{dL(A)}{da_i} = 2d_i a_i + 2(a_i - a_i^{(0)}) - \lambda = 0,$$

$$a_i = \frac{\lambda + 2a_i^{(0)}}{2(d_i + 1)} = \frac{\lambda}{2} \frac{1}{d_i + 1} + \frac{a_i^{(0)}}{d_i + 1}, \quad i=1, 2, \dots, p. \tag{8}$$

Substituting (8) into (7), find $\frac{\lambda}{2}$, we have:

$$\begin{aligned} \sum_{i=1}^p a_i &= \frac{\lambda}{2} \sum_{i=1}^p \frac{1}{d_i + 1} + \sum_{i=1}^p \frac{a_i^{(0)}}{d_i + 1} = 1 - a_0^{(0)}, \\ \frac{\lambda}{2} &= \frac{1 - a_0^{(0)} - \sum_{i=1}^p \frac{a_i^{(0)}}{d_i + 1}}{\sum_{i=1}^p \frac{1}{d_i + 1}}. \end{aligned} \tag{9}$$

Now, considering (8), (9), we will obtain:

$$a_i = \frac{(1 - a_0) - \sum_{i=1}^p \frac{a_i^{(0)}}{d_i + 1}}{\left(\sum_{i=1}^p \frac{1}{d_i + 1} \right) d_i + 1} + \frac{a_i^{(0)}}{d_i + 1}, \quad i=1, 2, \dots, p.$$

Thus, in the case that is important for practice when fuzzy source data are described by general functions of the (L-R) type, the analytic decision in the form of calculation formulas was obtained.

6. Discussion of results obtained in the modernization of methods for the identification of object states

Classic methods have a number of drawbacks, such as the lack of theoretical substantiation of the selection of the identification method; rigidity of the mechanism of conversion of the source data into the end result of identification; lack of informative value of results, which permits ambiguity of their interpretation.

Canonical technologies are based on the use of the theoretic-probabilistic description of the results of direct measurement of the controlled indicators of an object and the resulting indicators, assessing effectiveness of its functioning. The formed belief in the need to improve the approach to describing actual uncertainty of the source data led to the use for these purposes of the models and methods of the fuzzy sets theory and solution to the corresponding problems.

The modernization of the classic methods for solving the problem of identification of the state taking into consideration the fuzzy nature of the reproduction of the source data. In this case, we obtained analytical ratios describing the procedure for getting the ultimate results in a particular case that is important for practice when fuzzy source data are described for the functions of the (L-R) type.

The advantage of the proposed identification methods under conditions of uncertainty in comparison with the classic methods is explained by the possibility of solving this problem under actual conditions of a small sample of fuzzy source data. In this case, the proposed methods under uncertainty conditions can be adapted to solving the identification problems for any types of membership functions.

The identification methods proposed in research make it possible to reduce the identification error at a small sample of

fuzzy source data. In this case, these solutions can be applied in technologies using multidimensional discriminant analysis, clustering or regression analysis.

The major limitations when using the suggested methods include:

- the developed procedure does not ensure the adequate solution of the identification problem in case when source data are qualitative;
- the procedure is focused on the description of fuzzy source data by membership functions of the ($L-R$) type, while using membership functions of another type, the procedure of solving the problem gets complicated.

Further research into technology of the problems of identification of the state of objects in the context of fuzzy input data can be performed in the following areas:

- 1) improvement of the method of solving fuzzy systems of linear algebraic equations;
- 2) development of the methods for fuzzy optimization [28];
- 3) studying the results of application of the developed method for membership functions of different types.

7. Conclusions

1. The fuzzy method for discriminant analysis, which is the development of the corresponding classical method, was

proposed. It was established that when using it, it becomes possible to identify objects under the actual conditions of a small sample of fuzzy source data. Identification is possible by increasing the adequacy of descriptions of uncertain input data when constructing the discriminating surface. It should be noted that under this approach, the refusal from the theoretical-probabilistic technology of the source data description is a principal issue.

2. It was established that when using refined procedure of grouping with the separation of the grouping center, clustering becomes possible even at a small sample of fuzzy source data. This refined procedure provides for the use of the developed method of fuzzy clustering. Unlike similar methods, the original problem of comparison of fuzzy triangular numbers is reduced to the simpler problem of comparing a fuzzy number with zero.

3. It was established that when using the improved procedure for solving regression equations, it has become possible to assess regression coefficients analytically, if the source data are represented by general membership functions of the ($L-R$)-type. This improved procedure involves the use of the developed method of fuzzy regression analysis. In this case, under conditions of a small sample of fuzzy data, the adequacy of regression models improves due to taking into consideration the differences in the description of exogenous and endogenous variables.

References

1. The time course of individual face recognition: A pattern analysis of ERP signals / Nemrodov D., Niemeier M., Mok J. N. Y., Nestor A. // *NeuroImage*. 2016. Vol. 132. P. 469–476. doi: <https://doi.org/10.1016/j.neuroimage.2016.03.006>
2. Li D.-F. Multiattribute decision making models and methods using intuitionistic fuzzy sets // *Journal of Computer and System Sciences*. 2005. Vol. 70, Issue 1. P. 73–85. Doi: <https://doi.org/10.1016/j.jcss.2004.06.002>
3. Duda R., Hart P., Stork D. *Pattern Classification*. Wiley-Interscience, 2000. 688 p.
4. Borovikov V. P. *Iskusstvo analiza dannyh*. Piter: Sankt-Peterburg, 2005. 432 p.
5. Goia A., Vieu P. An introduction to recent advances in high/infinite dimensional statistics // *Journal of Multivariate Analysis*. 2016. Vol. 146. P. 1–6. doi: <https://doi.org/10.1016/j.jmva.2015.12.001>
6. Fisher discriminant analysis with kernels / Mika S., Ratsch G., Weston J., Scholkopf B., Mullers K. R. // *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)*. 1999. doi: <https://doi.org/10.1109/nnspp.1999.788121>
7. Bessokirnaya G. P. Diskriminantniy analiz dlya otbora informativnyh peremennyh // *Sociologiya: metodologiya, metody, matematicheskoe modelirovanie (4M)*. 2003. Issue 16. P. 25–35.
8. Bitjukov V. K. Formatirovanie klassov ob'ektov metodom diskriminantnogo mnogomernogo analiza // *Vestnik Voronezhskogo gosudarstvennogo universiteta inzhenernyh tekhnologiy*. 2001. Issue 6. P. 13–19.
9. Muhamediev B. M. *Ekonometrika i ekonometricheskoe prognozirovanie*. Almaty, 2007. 198 p.
10. Egorenko M. V., Bohovko A. G. Cluster analysis as a tool for grouping researched variablese // *Mezhdunarodniy nauchno-issledovatel'skiy zhurnal*. 2016. Issue 7. P. 25–29. doi: <http://doi.org/10.18454/IRJ.2016.49.096>
11. Hong Y., Kwong S. To combine steady-state genetic algorithm and ensemble learning for data clustering // *Pattern Recognition Letters*. 2008. Vol. 29, Issue 9. P. 1416–1423. doi: <https://doi.org/10.1016/j.patrec.2008.02.017>
12. Streke A., Ghosh J. Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions // *Journal of Machine Learning Research*. 2002. Issue 3. P. 583–617.
13. Krishna K., Narasimha Murty M. Genetic K-means algorithm // *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*. 1999. Vol. 29, Issue 3. P. 433–439. doi: <https://doi.org/10.1109/3477.764879>
14. Chang Y.-H. O., Ayyub B. M. Fuzzy regression methods – a comparative assessment // *Fuzzy Sets and Systems*. 2001. Vol. 119, Issue 2. P. 187–203. doi: [https://doi.org/10.1016/s0165-0114\(99\)00091-3](https://doi.org/10.1016/s0165-0114(99)00091-3)
15. Hong D. H., Lee S., Do H. Y. Fuzzy linear regression analysis for fuzzy input–output data using shape-preserving operations // *Fuzzy Sets and Systems*. 2001. Vol. 122, Issue 3. P. 513–526. doi: [https://doi.org/10.1016/s0165-0114\(00\)00003-8](https://doi.org/10.1016/s0165-0114(00)00003-8)

16. Sira O. V., Al-Shqeerat K. H. A New Approach for Resolving Equations with Fuzzy Parameters // *European Journal of Scientific Research*. 2009. Vol. 38, Issue 4. P. 619–625.
17. Optimizing h value for fuzzy linear regression with asymmetric triangular fuzzy coefficients / Chen F., Chen Y., Zhou J., Liu Y. // *Engineering Applications of Artificial Intelligence*. 2016. Vol. 47. P. 16–24. doi: <https://doi.org/10.1016/j.engappai.2015.02.011>
18. Yang M.-S., Lin T.-S. Fuzzy least-squares linear regression analysis for fuzzy input–output data // *Fuzzy Sets and Systems*. 2002. Vol. 126, Issue 3. P. 389–399. doi: [https://doi.org/10.1016/s0165-0114\(01\)00066-5](https://doi.org/10.1016/s0165-0114(01)00066-5)
19. Zack Y. A. Fuzzy-regression models under conditions of the presence of non-numeric data in the statistical sample // *System Research & Information Technologies*. 2017. Issue 1. P. 88–96. doi: <https://doi.org/10.20535/srit.2308-8893.2017.1.07>
20. Fuzzy pattern recognition-based approach to biometric score fusion problem / Fakhar K., El Aroussi M., Saidi M. N., Aboutajdine D. // *Fuzzy Sets and Systems*. 2016. Vol. 305. P. 149–159. doi: <https://doi.org/10.1016/j.fss.2016.05.005>
21. Raskin L. G., Seraya O. V. *Nechetkaya matematika. Osnovy teorii. Prilozheniya*. Kharkiv: Parus, 2008. 352 p.
22. Semenov S. G., Gavrylenko S. Y., Chelak V. V. Developing parametrical criterion for registering abnormal behavior in computer and telecommunication systems on the basis of economic tests // *Actual Problems of Economics*. 2016. Issue 4. P. 451–459.
23. Seraya O. V., Demin D. A. Linear regression analysis of a small sample of fuzzy input data // *Journal of Automation and Information Sciences*. 2012. Vol. 44, Issue 7. P. 34–48. doi: <https://doi.org/10.1615/jautomatinfscien.v44.i7.40>
24. Tymchuk S. Definition of information uncertainty in power engineering // *Technology audit and production reserves*. 2013. Vol. 6, Issue 5 (14). P. 33–35. URL: <http://journals.urau.ua/tarp/article/view/19648/17296>
25. Semenov S., Sira O., Kuchuk N. Development of graphicanalytical models for the software security testing algorithm // *Eastern-European Journal of Enterprise Technologies*. 2018. Vol. 2, Issue 4 (92). P. 39–46. doi: <https://doi.org/10.15587/1729-4061.2018.127210>
26. Multiservice network security metric / Mozhaev O., Kuchuk H., Kuchuk N., Mozhaev M., Lohvynenko M. // *2017 2nd International Conference on Advanced Information and Communication Technologies (AICT)*. doi: <https://doi.org/10.1109/aiact.2017.8020083>
27. Raskin L., Sira O. Method of solving fuzzy problems of mathematical programming // *Eastern-European Journal of Enterprise Technologies*. 2016. Vol. 5, Issue 4 (83). P. 23–28. doi: <https://doi.org/10.15587/1729-4061.2016.81292>
28. Strizhov V. V., Krymova E. A. *Metody vybora regressiionnyh modeley*. Moscow: VC RAN, 2010. 60 p.