DOI: http://doi.org/10.25728/cs.2023.6.5

A RANK-EXPERT DEVIATION FUNCTION TO CLASSIFY COMPLEX OBJECTS[1]

V.B. Korobov[2], A.G. Tutygin[3], and A.S. Lokhov[2]

[2]Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
[3]Laverov Federal Center for Integrated Arctic Research, Ural Branch, Russian Academy of Sciences, Arkhangelsk, Russia

✉ szoioran@mail.ru, ✉ andgt64@yandex.ru, ✉ a.s.lohov@yandex.ru

Abstract. This paper proposes a novel function for classifying environmental, social, and socio-environmental objects. It is based on the sum of rank deviations between a given object and a reference object considering the significance of the object's characteristics (factors). Characteristics are estimated using weight coefficients, which are provided by expertise or another method. A verbal numerical scale is developed to assess the proximity of objects by the numerical value of the deviation function. As is demonstrated below, this function is not a metric in the geometric sense but a proximity function defined in multidimensional scaling theory. As illustrative examples, the values of the deviation function are calculated for two applications: an environmental problem of comparing the vulnerability of territories to accidental oil spills and an economic problem of choosing real estate objects to purchase. A recommended sequence with a set of procedures based on the deviation function is presented to solve these problems.

Keywords: rank, object, classification, verbal numerical scale.

INTRODUCTION

Classification is one of the main problems of science. The goal of classification is to arrange objects so that those belonging to the same group can be considered close in their qualities. The result of classification is always a grouping of objects according to their properties.

The methodology of classifications varies from one science to another. We can mark off social and natural sciences, where a unified approach is difficult to develop. For example, researchers distinguish from 8 to 22 types of civilizations [1]. It is even harder to classify mixed objects consisting of social, natural, anthropogenic, and other components, e.g., different types of settlements. Nevertheless, mixed objects can be classified by comparing their components in some commensurable terms, such as scores, ranks, distances in hyperspaces, etc.

There exist no unified universal methods of classification. But they are actively developed by representatives of various sciences. Carl Linnaeus contributed much to developing the classifications of natural objects [2].

1. THE IDEOLOGY OF OBJECT COMPARISON

The methodology proposed below involves the component-wise comparison of objects. Their belonging to the same class is determined by a function, i.e., the total value of the deviations between the characteristics (indicators) of the corresponding components. This idea is not novel and is adopted in many applications. One of the simplest and most efficient functions is the well-known metric of Richard Hamming [3], originally developed by him within coding theory. Later, it found application in many fields of science and technology. In addition, there are several modifications, e.g., the weighted Hamming metric [4].

Consider the Hamming metric in detail. It has the general form

$$d\left(x_i,\ x_j\right)=\sum_{i,\,j=1}^{N}\left|x_i-x_j\right|,$$

where x_i and x_j are the coordinates of the vectors of

[1] This work was performed within state order no. FMWE-2021-0006 "Modern and ancient bottom sediments and suspended sediments of the World Ocean—a geological record of environmental and climate changes: dispersed sedimentary matter and bottom sediments of the seas of Russia, the Atlantic, Pacific and Arctic Oceans—lithological, geochemical, and micropaleontological studies; the research of pollution, paleoenvironments, and processes of marginal filters of rivers."

objects under comparison, $i, j = \overline{1, N}$; N denotes the number of object characteristics. The simplest method for determining the proximity of objects is to represent differences as binary relations: if the difference between the corresponding characteristics is below some threshold (e.g., 10%), it takes value 1, and value 0 otherwise. Next, the unities are summed up; if the result equals or exceeds some percentage of the total number of object characteristics (e.g., 90%), such objects are considered close and belong to the same class. The criteria of proximity and belonging (in the current example, 10% and 90%, respectively) are set by researchers arbitrarily depending on the conditions of a particular problem. There are no unified recommendations here, and it seems impossible to provide them. Meanwhile, the approach is quite simple.

Similar ideas are also employed in multidimensional scaling methods to analyze data by reducing their dimension, implemented by comparing objects in different ways. They are widely used in sociology and psychology, and their founder is L. Guttman [5, 6].

Currently, due to advances in machine learning, artificial intelligence, and cybernetics, the clustering problem [7] occupies a special place in science. This problem is similar to classification, in which the classes must be determined by an algorithm.

Nevertheless, various classification methods are still being developed. Among them, we note ATOVIC (Amended fused TOPSIS-VIKOR for classification) [8], a combination of two multicriteria choice methods modified for classification. In this method, an ideal object and the so-called negative object in terms of class belonging are determined for each class. Next, the proximity of each ideal object and each negative object to each class is checked using the Minkowski and Chebyshev metrics, and the belonging of each object is decided accordingly [9]. There exist other metric methods as well; *k* nearest neighbors [10] is the most popular algorithm. It has numerous modifications and can be used with different metrics.

In addition, we mention the naive Bayes classifier, which is based on the Bayes theorem. This classifier uses various probability statistics for decision-making [11].

The practical application of the Hamming metric (including our experience) demonstrated its effectiveness. At the same time, several disadvantages were revealed [12] (a common situation when a methodology developed for solving particular problems is translated to other objects). These disadvantages appear in other metric methods as well; the main one is that even a very high percentage of the matched pairs of characteristics with an equally high criterion of single pair matching (which can be set even below 5%) does not guarantee the belonging of objects to the same class.

For example, let some territories be geographically and ecologically classified by the degree of their pollution. Such an object is often described by selecting a finite number of most typical indicators (characteristics). In this case, almost all indicators, except for one, may have very close results (even completely coincide), giving all formal grounds to assign the same class to the objects under comparison. Nevertheless, a single mismatch may be so large (by orders of magnitude) that the objects cannot belong to the same class under any conditions. This situation occurs for accidental spills and large toxic releases causing environmental disasters. In such cases, the Hamming metric becomes inappropriate because it leads to erroneous results. This metric and its analogs have other drawbacks; see the presentation below.

Despite the considerations above, the principle of comparing and classifying objects based on the analysis of the deviations of their characteristics seems to be justified and methodologically correct. Based on this principle, we propose a function to compare and classify a wide range of objects.

2. THE RANK-EXPERT DEVIATION FUNCTION

To determine whether a given object belongs to a particular class, it must be compared with another object from this class. For such comparisons, we introduce the notion of a reference object, interpreted as an object characterizing a given class to the greatest extent. (How should a reference object be formalized? This issue will be discussed in a separate publication, as it requires a special methodology depending on the object's complexity and the availability of information.) From the viewpoint of computational procedures, a reference is also an object; therefore, we will consider a more general case, i.e., the comparison of two objects.

Deviations can be measured by several methods. The most natural approach is a simple difference expressed in percentage or Boolean symbols. Before that, it is possible to convert all characteristics into points using special scales and obtain differences in unified units. Indicators are often normalized to bring them to a dimensionless form. Other methods are also used to make object characteristics comparable.

This paper proposes a method based on ranking scales with one important property: verbal definitions have no sign, and negative and positive values, both quantitative and qualitative, can be therefore treated as unambiguous. This property is crucial for some characteristics. For example, when assessing the climate comfort for living, air temperature is estimated approximately in the range from –50°C to +40°C; when assessing the efficiency of industrial enterprises by their financial result, which can be both negative and positive, the degree of their unprofitability or profitability is estimated.

We now proceed to the function proposed in this paper. Consider two objects x and y with the sets of characteristics $\bar{\mathbf{x}} = \left(x_1,\ldots,\ x_N\right)$ and $\bar{\mathbf{y}} = \left(y_1,\ldots,\ y_N\right)$, respectively, where $\bar{\mathbf{x}},\ \bar{\mathbf{y}} \in \mathbb{R}^N$. In other words, the characteristics belong to the set of N–dimensional vectors, and their elements belong to the set of real numbers. Let the characteristics be represented in their "natural" units of measurement (e.g., physical). Then the vector of their deviations, $\bar{\mathbf{z}} \in \mathbb{R}_+^N \cup \left\{\mathbf{0}\right\}$, with nonnegative coordinates has the form

$$\bar{\mathbf{z}} = \left|\bar{\mathbf{x}} - \bar{\mathbf{y}}\right| = \left(z_1,\ldots,\ z_N\right). \tag{1}$$

However, such a vector quantity is not very informative and can only indicate the proximity of the values of objects' characteristics. For example, assume that the values of one characteristic of objects differ by 0.01 units, all others being equal. Does it mean that the objects are close to each other? Not necessarily; information about only two values is insufficient to assess the nature and range of variation of any characteristic.

One method for solving this problem is ranking: the values of characteristics are arranged in a certain order (in our case, by increasing the deviation of their values). Both quantitative and qualitative characteristics are subject to ranking, and this procedure can be therefore used to measure deviations by the difference in ranks between the objects under comparison. In this case, it is possible to obtain a quantitative estimate and also a verbal assessment of the degree of deviation for each characteristic ("small," "considerable," etc.), i.e., to determine its significance. It can be useful in some particular optimization problems, e.g., in the analysis of alternatives.

This approach requires developing special verbal numerical ranking scales for each characteristic. However, the corresponding costs are covered by the possibility of their independent application in other problems.

We introduce a nonnegative matrix of private ranking scales, $\mathbf{H} \in \mathbb{M}_{N\times M}\left(\mathbb{R}_+ \cup \left\{0\right\}\right)$. In this matrix, row $i = \overline{1,\ N}$ corresponds to the scale of the characteristic with the same number:

$$\mathbf{H} = \begin{pmatrix} h_{1,1} & \cdots & h_{1,M} \\ \vdots & \ddots & \vdots \\ h_{N,1} & \cdots & h_{N,M} \end{pmatrix},$$

where $M \in \mathbb{N}$ is the maximum rank value, $M \geq 1$. Then the number of gradations or intervals of the ranking scale equals $(M+1)$. (For $M = 1$, we obtain the binary scale.) According to the aforesaid, the scale values must be arranged in ascending order, from the left column of the matrix **H** to the right one: for each i, $0 \leq h_{i,1} < \ldots < h_{i,M}$. Different methods are used to construct ranking scales. Based on our experience, it seems preferable to design them by an expert survey within the theory of fuzzy sets.

Different objects can be described by a different number of characteristics; moreover, the same object may have different lengths of components depending on its level of scrutiny or the conditions imposed on the accuracy of its classification. Therefore, the following question arises naturally: what shall we do if the number of gradations is not the same for each component? In this case, a technically simple scale synchronization procedure can be proposed: scale factors are introduced to bring all scales to the same number of gradations. But such situations should be quite justifiably avoided, and objects are often described using scales with the same number of gradations. For example, a single verbal scale of five gradations was developed for all seven characteristics of the atmosphere [13] to classify the atmospheric pollution potential. This verbal scale has been used in practice for several decades.

We define the vector of rank deviations $\bar{\mathbf{r}} = \left(r_1,\ldots,\ r_N\right)$ of objects x and y, $\bar{\mathbf{r}} \in \mathbb{N}_0^N$, such that $0 \leq r_i \leq M,\ i = \overline{1,\ N}$. Then

$$r_i = \sum_{j=1}^{M} \theta(z_i - h_{i,j}), \tag{2}$$

where θ denotes the Heaviside function:

$$\theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0, \end{cases} \quad x \in \mathbb{R}.$$

For each i, formula (2) yields the number of values $h_{i,j}$ strictly less than z_i. This number is the deviation rank, which varies from 0 to M inclusive, where M is the number of columns in the matrix H and $(M + 1)$ is the number of gradations of private ranking scales. This expression can also be written in vector form when replacing the Heaviside function with the function of counting all nonnegative elements of the vector.

All object's characteristics are unequal by their role in the object's functioning or intended purpose. This value inequality of characteristics is considered through weight coefficients expressed in unit fractions or percentage. We introduce the vector $\bar{\mathbf{k}} = \left(k_1,\ldots,k_N\right)$, $\bar{\mathbf{k}} \in \mathbb{R}_+^N \cup \left\{\mathbf{0}\right\}$, of weight coefficients such that $\sum_{i=1}^{N} k_i = 1$.

Then the rank-expert deviation function is defined as a mapping from the Cartesian product of the two sets of N-dimensional nonnegative real vectors into the set of nonnegative real numbers, i.e., $R : \left(\mathbb{R}_+^N \cup \left\{\mathbf{0}\right\}\right) \times \left(\mathbb{R}_+^N \cup \left\{\mathbf{0}\right\}\right) \to \mathbb{R}_+ \cup \left\{0\right\}$ and is given by

$$R\left(\bar{\mathbf{k}},\ \bar{\mathbf{r}}\right) = \frac{\bar{\mathbf{k}} \cdot \bar{\mathbf{r}}}{M} = \sum_{i=1}^{N} \frac{k_i r_i}{M}, \tag{3}$$

where $\cdot$ denotes the inner product of vectors. Note that in formula (3), the vector of rank deviations $\bar{\mathbf{r}}$ is also normalized by the maximum rank value. Therefore, the range of R is the set of unit fractions, which can be represented in percentage terms for convenience.

The function described above is not a metric: it satisfies neither the identity axiom nor the triangle inequality. Therefore, this function does not define the distance between objects in some space in the conventional geometric sense. Within the theory of multidimensional scaling, this problem is solved using proximity functions defined as follows [6, *p. 39*]. Consider three vectors of object's characteristics, a_i, a_j, and a_k. A proximity function $s\left(a_i, a_j\right)$ is a function such that, for all i, j, k, the following relations (axioms) are satisfied: $s(a_i,\ a_i) \geq s(a_i,\ a_j)$; $s(a_i,\ a_j) = s(a_j,\ a_i)$; for large values $s\left(a_i, a_j\right)$ and $s\left(a_j, a_k\right)$, the value $s\left(a_i, a_k\right)$ has at least the same order [6, *p. 39*].

These relations are the weakened axioms from the definition of a metric. As is easily checked, they hold for the rank-expert deviation function defined above, so it represents a proximity function. In particular, the affine transformation, often used in multidimensional scaling methods for scaling and dimension reduction, can be applied to this function as well.

One of the main advantages of the rank-expert deviation function is the possibility to establish equality criteria for values. For example, the objects whose measured physical quantities differ at most by the measurement error should be assumed equal. For such objects, the first gradation of the private ranking scale (the first column of the matrix H) should be not less than the measurement error. In the general case, however, it can be 0, meaning the absence of deviation only if the characteristics are exactly equal. This situation applies, e.g., to qualitative variables.

In this context, the following questions seem natural. Why is it necessary to use private ranking scales? Is it prohibited to take the ratio of characteristics in percentage? For example, formulas (1) and (2) can be in theory replaced by

$$z_i = 1 - \frac{\min(x_i,\ y_i)}{\max(x_i,\ y_i)},\ \ i = \overline{1,\ N}. \tag{4}$$

We provide the answer with an illustrative example. Let objects be some water masses characterized by the concentration of suspended solids in water, measured in mg/l. It has a rather wide range of variation, taking values in a few mg/l and in a few hundredths of mg/l. If two objects under comparison have concentrations of 0.01 mg/l and 0.02 mg/l, respectively, their difference by formula (4) is 50%; for objects with values of 2 mg/l and 0.01 mg/l, respectively, it equals 99.5%. In this case, the objects with values of 0.01 mg/l and 0.02 mg/l may differ insignificantly within a particular problem or can be taken as equal under a measurement accuracy of 0.01 mg/l. Therefore, in this example, the 50% difference in characteristics becomes incorrect.

3. A UNIVERSAL CLASSIFICATION SCALE FOR OBJECTS

A special scale is needed to assess the proximity of objects under comparison. We propose a possible solution based on the following considerations. According to the aforesaid, the function achieves maximum for the largest difference between the objects, i.e., when the deviation ranks of each characteristic take their maximum value M. The minimum value of the function is 0. (In this case, one object strictly matches another, and all deviations are 0.) Consequently, all possible deviations lie in the range from 0 to 1, or 100%. As a result, the following estimation scale can be constructed (Table 1).

Other modifications of this scale are possible, but it is a matter of discussion. (Note that this scale should be designed by fuzzy set methods.) The graphical form of this scale is presented in Fig. 1. The private ranking scales making up the matrix H have the same form as well.

Table 1

The universal scale for assessing the proximity of objects by the rank-expert deviation function

No.	Interval, %	The degree of proximity
1	0–5	Coincidence
2	6–10	Minor discrepancy
3	11–20	Slight discrepancy
4	21–30	Moderate discrepancy
5	31–40	Notable discrepancy
6	41–55	Substantial discrepancy
7	56–70	Significant discrepancy
8	71–100	Very significant discrepancy

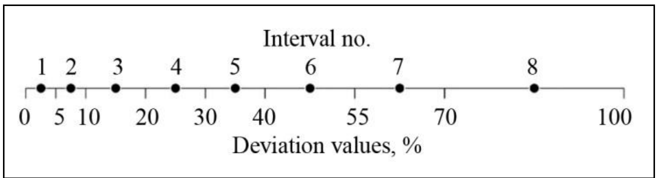

Fig. 1. The universal scale for assessing the proximity of objects. Numbers above the scale indicate the intervals according to Table 1.

4. NUMERICAL EXAMPLES

Example 1. This example primarily demonstrates the technical aspects of applying the calculation procedures: the techniques for selecting the private ranking scales and weight coefficients are beyond the scope of this paper.

Consider an environmental problem of comparing the vulnerability of territories to accidental oil spills [14]. In this case, the objects are the domains with the available values of indicators. We adopt three indicators:

• the forecasted (simulation-based) pollution area for a particular sector of the terrain (m^2), used to assess the degree of damage and the scope of restoration works;

• the average surface slope (%), affecting the rate of pollution spreading on the ground surface and the shape of the oil spot;

• the share of waterbodies (%), where oil and oil products are transported by water streams to significant distances, thereby increasing the pollution area (banks and coastal territories) and environmental damage.

For these indicators, private ranking scales with four gradations were developed and weight coefficients were determined. Thus, N = 3 and M = 4; the input data for the calculations are provided in Table 2.

Note that all values in this table, except for columns 1 and 4, are given in the indicator units. Columns 5–8 of Table 2 present the matrix of private ranking scales H. Next, Table 3 shows the vector of the deviations of object's characteristics (formula (1), column 2), its transformation into the vector of rank deviations (formula (2), columns 3–6 and 7), and the scalar product (formula (3), column 8). The transformation into the ranking scales is carried out by counting positive values in columns 3–6, which is the rank value (column 7).

The value of the rank-expert deviation function is 0.725; it means that the objects differ from each other by 72.5%. According to the universal classification scale, it corresponds to a very significant discrepancy by the degree of proximity (see Table 1).

Table 2

Input data for calculating the rank-expert deviation function

Characteristic	i	$\overline{x}$	$\overline{y}$	$\overline{k}$	$h_{i,1}$	$h_{i,2}$	$h_{i,3}$	$h_{i,4}$
Column no.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pollution area, m^2	1	2376.7	1831.6	0.5	10	25	100	200
Surface slope, %	2	1.05	1.13	0.3	0.05	0.25	0.5	1
The share of waterbodies, %	3	2	0.24	0.2	0.01	0.1	1	5

Table 3

Calculation results for the rank-expert deviation function

i	$\overline{z}$	$z_i - h_{i,1}$	$z_i - h_{i,2}$	$z_i - h_{i,3}$	$z_i - h_{i,4}$	r_i	$r_i \cdot k_i / M$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	545.1	535.1	520.1	445.1	345.1	4	0.5
2	0.08	0.03	–0.17	–0.42	–0.92	1	0.075
3	1.76	1.75	1.66	0.76	–3.24	3	0.15
The value of the rank-expert deviation function							R = 0.725

Such tables can be constructed for each sector of the terrain of the object or a separate alternative. Then we obtain some set of objects for classification, which is a primary goal when studying objects and alternatives, e.g., in industrial safety problems and choice problems (see below). ♦

Example 2. As another example, we solve a classical economic problem of choosing real estate objects (apartments in a building). There are three alternatives (*a*, *b*, and *c*) and five different characteristics of apartments. We denote their vectors by $\bar{\mathbf{a}}$, $\bar{\mathbf{b}}$, and $\bar{\mathbf{c}}$ (Table 4, columns 3–5). The easiest way to form a reference object ($\bar{\mathbf{e}}$) is to take the minimum or maximum possible values of the characteristics. For example, for the depreciation of the building (in %), the acceptable value of the reference is 0: the less the depreciation is, the better the alternative will be (Table 4, column 2). We form the vector of weight coefficients (the fastest method is simple ranking) and arrange the characteristics in descending order of their significance by number *i*: 1, 2, 3, 4, and 5. As a result, for this problem, $k_i = (6-i)/15$ (Table 4, column 6). Next, we construct the matrix of private ranking scales $\mathbf{H}$ with four gradations or intervals for each characteristic. Then $N = 5$ and $M = 3$. In this case, to simplify the example, the scales are formed by questioning one expert (the purchaser or his or her representative). There may be several experts; as we believe, the most appropriate number is about 10 people. Also, the scale can be designed by other methods. Thus, all the initial data necessary for the rank-expert deviation function have been formed (see Table 4).

Let us proceed to the calculations, i.e., the sequential application of this function to compare the reference with each alternative. For convenience, the first few steps of calculations are omitted; they are carried out by analogy with the previous example (see Table 3, columns 2–6). Table 5 shows the rank values (columns 2, 4, and 6), their normalization (columns 3–5), and the values of the rank-expert deviation function for each alternative.

In this case, the best alternative is *b* as the one with the smallest value of the rank-expert deviation function $R = 0.553$. (It differs least from the reference.) ♦

Table 4

Input data for calculating the rank-expert deviation function

Characteristic	i	$\bar{e}$	$\bar{a}$	$\bar{b}$	$\bar{c}$	$\bar{k}$	$h_{i,1}$	$h_{i,2}$	$h_{i,3}$
Column no.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Value, thousand rubles/m^2	1	70	80	100	150	0.33	5	10	50
Distance to the city center, km	2	1	15	5	6	0.27	5	10	15
The depreciation of the building, %	3	0	30	60	10	0.2	10	25	50
The number of parking lots near the building, pcs.	4	1000	500	50	100	0.07	100	500	1000
Apartment area, m^2	5	60	63	78	55	0.13	5	10	25

Table 5

Calculation results for the rank-expert deviation function

Alternative	i	$\bar{a}$		$\bar{b}$		$\bar{c}$	
		r_i	$r_i \cdot k_i / M$	r_i	$r_i \cdot k_i / M$	r_i	$r_i \cdot k_i / M$
Column no.	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Value, thousand rubles/m^2	1	2	0.220	2	0.220	3	0.330
Distance to the city center, km	2	2	0.180	0	0.000	1	0.090
The depreciation of the building, %	3	2	0.133	3	0.200	1	0.067
The number of parking lots near the building, pcs.	4	2	0.047	2	0.047	2	0.047
Apartment area, m^2	5	0	0.000	2	0.087	1	0.043
The value of the rank-expert deviation function R		0.580		0.553		0.577	

5. APPLICATIONS OF THE RANK-EXPERT DEVIATION FUNCTION

The function proposed above serves to classify objects with characteristics (indicators) expressed in a variety of quantitative and qualitative values. First of all, such objects include social and economic, geographical, and ecological objects. They are formalized by considering many factors from various fields of knowledge: natural, technical, social, military, geopolitical, and others. Ranking scales make such characteristics commensurable. It is possible to use available verbal numerical scales provided that they correspond to the problem conditions.

Generally speaking, the rank-expert deviation function and its components can be treated as one element of classification technologies [15] and can be used for such complex problems as the geographical and ecological zoning of territories [14] and others. We will demonstrate it on some intentionally simplified examples from real life.

Consider again the vulnerability of territories to accidental oil spills along a pipeline route (see Example 1 in Section 4). Let us divide the entire route into equally long sectors. The length and width of the sectors are of no fundamental importance: due to the capabilities of GIS technologies, they can be set as small as desired. The rank-expert deviation function calculated for each sector (see Table 4) is mapped (Fig. 2) to assess the vulnerability of the entire object, to develop scientifically grounded recommendations for industrial environmental monitoring, and to select optimal locations for emergency spill response equipment (barrier booms, sorbents, etc.). For example, such equipment should be placed in the sector of the terrain with the highest vulnerability (according to Fig. 2, in the northeastern sector with $R = 0.87$).

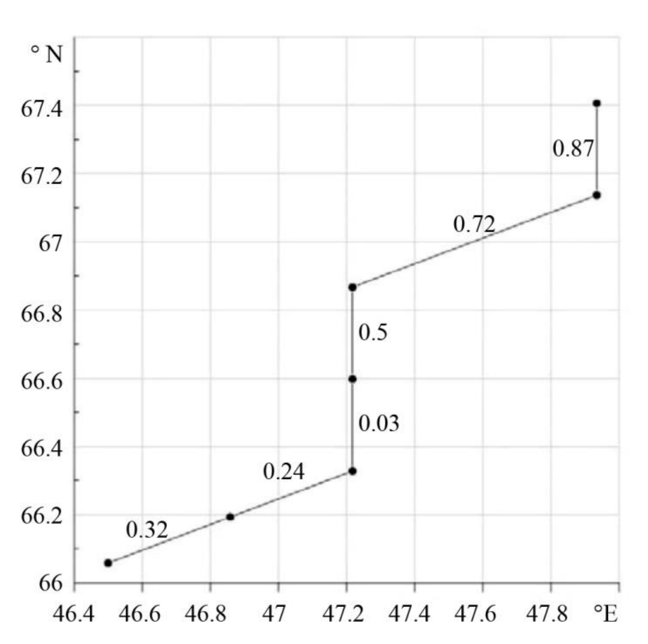

Fig. 2. The value diagram of the rank-expert deviation function for the fragment of a linear object (oil pipeline).

CONCLUSIONS

An object classification methodology based on the rank-expert deviation function has been proposed. It represents a set of procedures performed in the following recommended sequence.

1) The goal of classification is determined.

2) According to this goal, object's characteristics are selected for an appropriate formalization of the object; the number of such characteristics is not limited but must be the same for all objects.

3) Private verbal numerical ranking scales are constructed (if available, selected) for each characteristic of the object. All of them must have the same number of gradations or be preliminarily synchronized using a correction factor.

4) A reference object with the most appropriate characteristics is designed for each class.

5) The weight coefficients of the characteristics are calculated.

6) After determining the initial data, the rank-expert deviation function is calculated in three steps.

6.1) The vector of the deviations of characteristics is constructed.

6.2) This vector is transformed into the vector of rank deviations.

6.3) The vector of rank deviations is normalized, and the inner product of the normalized one with the vector of weight coefficients is obtained.

7) The verbal numerical scale is used to find the degree of proximity of the object under comparison and the reference objects; if necessary, a sequential comparison with other reference objects is made.

This function may serve to solve other problems, e.g., simple comparison of two objects with specified characteristics in the analysis of alternatives, scenario design problems, and object optimization.

However, all values of the rank-expert deviation function may fall into the same gradation, i.e., the object is homogeneous. This situation sometimes occurs in practice. In such cases, additional classification rules are developed to differentiate the object under consideration. It often suffices to add one or two conditions; e.g., among equally important alternatives, priority can be given to the one where the greatest contribution is made by the characteristic with the highest weight factor and the minimum deviation from the reference.

REFERENCES

1. Huntington, S.P., *The Clash of Civilizations and the Remaking of World Order*, New York: Simon & Schuster, 2011.
2. Linnaeus, C., *Linnaeus' Philosophia Botanica*, Oxford: Oxford University Press, 2005.
3. Hamming, R.W., Error-Detecting and Error-Correcting Codes, *Bell System Technical Journal*, 1950, vol. 29, no. 2, pp. 147–160.
4. Deza, M. and Deza, E., *Encyclopedia of Distances*, Berlin–Heidelberg: Springer-Verlag, 2009.
5. Guttman, L., A General Nonmetric Technique for Finding the Smallest Coordinate Space for a Configuration of Points, *Psychometrika*, 1968, vol. 23, no. 4, pp. 469–506.
6. Tolstova, Yu.N., *Osnovy mnogomernogo shkalirovaniya* (Foundations of Multidimensional Scaling), Moscow: KDU, 2006. (In Russian.)
7. Everitt, B.S., Landau, S., Leese, M., and Stahl, D., *Cluster Analysis*, 5th ed., Hoboken: John Wiley & Sons, 2011.DOI: 10.1002/9780470977811.
8. Baccour, L., Amended Fused TOPSIS-VIKOR for Classification (ATOVIC) Applied to Some UCI Data Sets, *Expert Systems with Applications*, 2018, vol. 99, pp. 115–125. DOI: 10.1016/j.eswa.2018.01.025.
9. Yusuf, H. and Panoutsos, G., Multi-criteria Decision Making Using Fuzzy Logic and ATOVIC with Application to Manufacturing, *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Glasgow, 2020, pp. 1–7. DOI: 10.1109/FUZZ48607.2020.9177772.
10. Taunk, K., De, S., Verma, S., and Swetapadma, A., A Brief Review of Nearest Neighbor Algorithm for Learning and Classification, *Proceedings of the International Conference on Intelligent Computing and Control Systems (ICCS)*, Madurai, 2019, pp. 1255–1260. DOI: 10.1109/ICCS45141.2019.9065747.
11. Yang, F.J., An Implementation of Naive Bayes Classifier, *Proceedings of the International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, 2018, pp. 301–306. DOI: 10.1109/CSCI46756.2018.00065.
12. Korobov, V.B. and Tutygin, A.G., *Klassifikatsionnye metody resheniya ekologo-ekonomicheskikh zadach* (Classification Methods for Solving Ecological and Economic Problems), Arkhangelsk: Pomorskii University, 2010. (In Russian.)
13. Bezuglaya, E.Yu., *Monitoring sostoyaniya atmosfery v gorodakh* (Atmosphere Monitoring in Cities), Leningrad: Gidrometeoizdat, 1986. (In Russian.)
14. Lokhov, A.S., Gubaidullin, M.G., Korobov, V.B, and Tutygin, A.G., Geographical and Ecological Land Zoning of Onshore Oil Pipeline Location by Level of Hazard to Environment from Emergency Oil Spills in Arctic Region, *Theoretical and Applied Ecology*, 2020, no. 4, pp. 43–48.
15. Korobov, V.B., Kochurov, B.I., and Tutygin, A.G. Methodology of Zoning of Complex Geographic and Ecological Objects Using Expert Statistical Methods, *Regional Environmental Issues*, 2020, no. 5, pp. 42–48. (In Russian.)

This paper was recommended for publication by F.T. Aleskerov, a member of the Editorial Board.

Received July 26, 2023,
and revised September 13, 2023.
Accepted September 28, 2023.

Author information

Korobov, Vladimir Borisovich. Dr. Sci. (Geo.), Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
✉ szoioran@mail.ru
ORCID iD: https://orcid.org/0000-0002-3198-9208

Tutygin, Andrei Gennad'evich. Cand. Sci. (Phys.–Math.), Laverov Federal Center for Integrated Arctic Research, Ural Branch, Russian Academy of Sciences, Arkhangelsk, Russia
✉ andgt64@yandex.ru
ORCID iD: https://orcid.org/0000-0001-9821-651X

Lokhov, Alexei Sergeevich. Cand. Sci. (Geo.), Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
✉ a.s.lohov@yandex.ru
ORCID iD: https://orcid.org/0000-0001-5022-9071

Cite this paper

Korobov, V.B., Tutygin, A.G., and Lokhov, A.S., A Rank-Expert Deviation Function to Classify Complex Objects. *Control Sciences* **6,** 48–55 (2023). http://doi.org/10.25728/cs.2023.6.5

Translated into English by *Alexander Yu. Mazurov*,
Cand. Sci. (Phys.–Math.),
Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
✉ alexander.mazurov08@gmail.com