

Indicators of Similarity and Dissimilarity of Multi-Attribute Objects in the Metric Spaces of Sets and Multisets

A. B. Petrovsky^{a, b, c, *}

^aFederal Research Center “Computer Science and Control”, Russian Academy of Sciences, Moscow, 119333 Russia

^bNational Research University Belgorod State University, Belgorod, 308015 Russia

^cShukhov Belgorod State Technological University, Belgorod, 308012 Russia

*e-mail: pab@isa.ru

Abstract—The concepts of the similarity and dissimilarity (difference) of analyzed objects play an important role in many theoretical and practical problems of decision making, artificial intelligence, pattern recognition, processing of heterogeneous information, etc. The similarity or dissimilarity of objects is usually estimated by their proximity in the attribute space. This paper considers new classes of metric spaces of finite, bounded, and measurable sets and multisets. The possibilities are shown for using new types of metrics to evaluate the similarity or dissimilarity of multi-attribute objects that are present in several instances with differing values of attributes and are represented by multisets of attributes.

Keywords: multi-attribute objects, similarity and dissimilarity of objects, multiset, metric spaces of sets and multisets, metrics, pseudometrics, quasi-metrics

DOI: 10.3103/S0147688218050052

INTRODUCTION

In various theoretical and practical problems of decision making, artificial intelligence, pattern recognition, data processing, and in other subject areas there is often a need to estimate the similarity or dissimilarity of studied objects (variants or alternatives) on the basis of their properties, which are expressed by attributes of objects. The similarity or dissimilarity of objects is usually estimated by their proximity in the attribute space.

There is a rather wide range of problems where the analyzed objects are given by many heterogeneous attributes, that is, quantitative, qualitative, mixed, continuous or discrete ones; as well, the same objects are present in several instances (versions or copies) with different values of attributes, whose convolution is either impossible or mathematically incorrect. Different versions of object description arise in cases when an object is estimated by several experts for many criteria with numerical and/or verbal scales, or the characteristics of the object are calculated several times by different methods, or measured several times by different instruments. Examples of such problems are group classification and group ordering of multi-attribute decisions, recognition of graphic symbols, and processing of text documents. The multiplicity and repetition of factors that describe objects complicates and hinders the solution of such problems.

A convenient mathematical model for representing the multi-attribute objects is a multiset or a set with

repeating elements. The multiplicity of elements is an essential attribute of the multiset, which distinguishes it from the set. This paper considers new types of metric spaces of sets and multisets. The possibilities of application of new types of pseudometrics and quasi-metrics for estimating the similarity or dissimilarity of multi-attribute objects that are present in several instances with differing attribute values and are represented by multisets of attributes are shown.

1. MULTISSETS AND OPERATIONS ON MULTISSETS

Let us briefly outline the main theses of the theory of multisets [6, 7]. A *multiset* A generated by an ordinary (crisp) set $X = \{x_1, x_2, \dots\}$, all elements x_i of which are distinct, is defined as a set of groups of identical elements written in the form

$$A = \{k_A(x) \circ x \mid x \in X, k_A(x) \in \mathbb{Z}_+\}. \quad (1)$$

Here, $k_A: X \rightarrow \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ is called a *function of the number of instances* of a multiset, which specifies the multiplicity of the occurrence of element $x \in X$ in multiset A that is denoted by symbol \circ . The multiset is a generalization of the concept of a set and becomes a set for $k_A(x) = \chi_A(x)$, where $\chi_A(x) = 1$ for $x \in A$ and $\chi_A(x) = 0$ for $x \notin A$. Group $k_A(x) \circ x = x^{k_A(x)}$ that unites k identical elements x is called a *component* of multiset A . A one-element set $\{x\} = \{x^1\} = \{1 \circ x\}$ is called a *simple*

set or a *singleton*, while a one-component multiset $\{x^k\} = \{k \circ x\}$ is called a *simple multiset*, *multiple singleton*, or *multiton*. A “zero-point” set $\{x^0\} = \{0 \circ x\}$ is called a *zeron*.

The set $\text{Supp}A = \{x \in A \mid \chi_{\text{Supp}A}(x) = \min[k_A(x), 1]\}$ is called a *carrier* or a *support set* of the multiset A . The *cardinality* of a multiset is defined as the total number of all instances of elements of multisets, equal to the sum of their multiplicities: $\text{card}A = |A| = \sum_x k_A(x)$; the *dimension* of a multiset is defined as the total number of single instances of all elements of a multiset, equal to the sum of their single multiplicities: $\text{dim}A = /A/ = \sum_x \chi_A(x) = |\text{Supp}A|$. The maximum value of the multiplicity function $\text{alt}A = \max_{x \in X} k_A(x)$ is called the *height* of multiset A .

Multisets A and B are said to be *equal* ($A = B$) when $k_A(x) = k_B(x), \forall x \in X$. Multisets A and B are *unequal* ($A \neq B$) when $k_A(x) \neq k_B(x)$ for at least one $x \in X$. For equal multisets, we have $|A| = |B|, /A/ = /B/, \text{Supp}A = \text{Supp}B, \text{alt}A = \text{alt}B$. Multisets A and B are said to be of *equal cardinality* if $|A| = |B|$; of *equal dimension* if $/A/ = /B/$; of *equal size* if they are of equal cardinality and equal dimension. Equal multisets are equal in size, the converse statement, generally speaking, is not true. The equality of multisets is an equivalence relation, since it is reflexive ($A = A$), symmetric ($A = B, B = A$), and transitive ($A = B, B = C \Rightarrow A = C$).

It is said that multiset A is included in multiset B ($A \subseteq B$) when $k_A(x) \leq k_B(x), \forall x \in X$. Multiset A is then called a *submultiset* or a *multisubset* of multiset B , and multiset B is an *overmultiset* or an *initial, parent* multiset of multiset A . In this case $|A| \leq |B|, /A/ \leq /B/, \text{Supp}A \subseteq \text{Supp}B, \text{alt}A \subseteq \text{alt}B$. As in the case of sets, the simultaneous fulfillment of conditions $A \subseteq B$ and $B \subseteq A$ implies the equality of multisets $A = B$. If $A \subseteq B$, but $B \not\subseteq A$, then multiset A is called the proper submultiset of multiset B and denoted as $A \subset B$. The inclusion of the multiset is the preorder relation, since it is reflexive ($A \subseteq A$) and transitive ($A \subseteq B, B \subseteq C \Rightarrow A \subseteq C$).

Multiset Z is called *maximal* if all multisets A of the family of multisets \mathbf{A} are submultisets of multiset Z ($A \subseteq Z$), that is, if $k_A(x) \leq k_Z(x)$; *empty* \emptyset if $k_{\emptyset}(x) = 0$; *constant* $N_{[h]}$ if $k_{N_{[h]}}(x) = h = \text{const}, h \in \mathbb{Z}_+, \forall x \in X$. Thus, the empty multiset \emptyset is a constant multiset $N_{[0]}$ of height 0, and the common set A , in particular, the support $\text{Supp}A$ of an arbitrary multiset A , is a constant

multiset $N_{[1]}$ of height 1. The maximal multiset Z can also be considered as a constant multiset $N_{[k]}$ of height $k = k_Z(x) = \max_{A \in \mathbf{A}} k_A(x), \forall x \in X$.

Let us assume that all the multisets that constitute family $\mathbf{A} = \{A_i\}_{i \in I}$ are submultisets of the same maximal multiset Z generated by some set X . Let us define the following operations on multisets:

union

$$A_1 \cup \dots \cup A_n = \bigcup_{i \in I} A_i = \{k_{\cup_i A_i}(x) \circ x \mid k_{\cup_i A_i}(x) = \max_{i \in I} k_{A_i}(x), \forall x \in X\};$$

intersection

$$A_1 \cap \dots \cap A_n = \bigcap_{i \in I} A_i = \{k_{\cap_i A_i}(x) \circ x \mid k_{\cap_i A_i}(x) = \min_{i \in I} k_{A_i}(x), \forall x \in X\};$$

addition

$$A_1 + \dots + A_n = \sum_{i \in I} A_i = \left\{ k_{\sum_i A_i}(x) \circ x \mid k_{\sum_i A_i}(x) = \sum_{i \in I} k_{A_i}(x), \forall x \in X \right\};$$

subtraction

$$A - B = \{k_{A-B}(x) \circ x \mid k_{A-B}(x) = \max[k_A(x) - k_B(x), 0], \forall x \in X\};$$

symmetric difference

$$A \Delta B = \{k_{A \Delta B}(x) \circ x \mid k_{A \Delta B}(x) = |k_A(x) - k_B(x)|, \forall x \in X\};$$

complement

$$\bar{A} = Z - A = \{k_{\bar{A}}(x) \circ x \mid k_{\bar{A}}(x) = k_Z(x) - k_A(x), \forall x \in X\};$$

multiplication by a scalar (reproduction)

$$b \cdot A = \{k_{b \cdot A}(x) \circ x \mid k_{b \cdot A}(x) = b \cdot k_A(x), b \in \mathbb{Z}_+, \forall x \in X\};$$

multiplication

$$A_1 \cdot \dots \cdot A_n = \prod_{i \in I} A_i = \left\{ k_{\prod_i A_i}(x) \circ x \mid k_{\prod_i A_i}(x) = \prod_{i \in I} k_{A_i}(x), \forall x \in X \right\};$$

n-th power

$$A^n = \{k_{A^n}(x) \circ x \mid k_{A^n}(x) = (k_A(x))^n, \forall x \in X\};$$

direct product

$$\begin{aligned} & A_1 \times \dots \times A_n \\ &= \left\{ k_{A_1 \times \dots \times A_n}(x_{p_1}, \dots, x_{p_n}) \circ \langle x_{p_1}, \dots, x_{p_n} \rangle \mid k_{A_1 \times \dots \times A_n}(x_{p_1}, \dots, x_{p_n}) \right. \\ & \quad \left. = \prod_i k_{A_i}(x_{p_i}), x_{p_i} \in A_i, i = 1, \dots, n \right\}; \end{aligned}$$

direct n -th power

$$(\times A)^n = \left\{ k_{(\times A)^n}(x) \circ x \mid k_{(\times A)^n}(x) = \prod_i k_A(x_i), x_i \in A, i = 1, \dots, n \right\}.$$

The supports of operations on multisets satisfy the following relations:

$$\begin{aligned} \text{Supp}(A \cup B) &= \text{Supp}(A + B) \\ &= (\text{Supp}A) \cup (\text{Supp}B); \end{aligned}$$

$$\begin{aligned} \text{Supp}(A \cap B) &= \text{Supp}(A \cdot B) \\ &= (\text{Supp}A) \cap (\text{Supp}B); \end{aligned}$$

$$\text{Supp}(A \Delta B) = (\text{Supp}(A - B)) \cup (\text{Supp}(B - A));$$

$$\text{Supp}(b \cdot A) = \text{Supp}A = \text{Supp}(A^n);$$

$$\text{Supp}(A \times B) = (\text{Supp}A) \times (\text{Supp}B);$$

$$\text{Supp}(\times A)^n = (\times \text{Supp}A)^n.$$

Many of the properties of operations on multisets are analogous to the properties of operations on ordinary sets, including the idempotency of union and intersection $A \cup \dots \cup A = A$, $A \cap \dots \cap A = A$, involution (double negation) of complementation $\overline{\overline{A}} = A$, identity, commutativity, associativity, and distributivity of some operations. As in the case of sets, not all operations on multisets are mutually commutative, associative, and distributive. In addition, operations of addition, multiplication by a scalar, multiplication, and raising to an arithmetic power are not defined for sets. In the transition from multisets to sets, operations of multiplication and raising to an arithmetic power turn into intersection, and operations of addition and multiplication by a scalar become impossible.

For multisets, as well as for sets, there is a duality of the operations of union and intersection in relation to the complementation operation, similar to the laws of De Morgan, and, in addition, there is a new type of duality of operations of arithmetic addition and subtraction:

$$\begin{aligned} \overline{A \cup B} &= \overline{A} \cap \overline{B}, \quad \overline{A \cap B} = \overline{A} \cup \overline{B}; \\ \overline{A + B} &= \overline{A} - B = \overline{B} - A, \\ \overline{A - B} &= \overline{A} + B, \quad \overline{A - B} = B - A. \end{aligned}$$

Multisets lack some of the properties of operations that sets have. At the same time, new properties appear that have no analogues for sets. As an example, in the theory of sets, the following equalities are always satisfied: $A \cap \overline{A} = \emptyset$, $A \cup \overline{A} = U$, where U is a universal set. Unlike sets, for multisets the following relations, which follow from the definition of complementation

of a multiset, are satisfied: $A + \overline{A} = Z$, $Z - \overline{A} = A$. At the same time, both $A \cap \overline{A} = \emptyset$, $A \cup \overline{A} = Z$, $A - \overline{A} = \overline{A} - A = \emptyset$ and $A \cap \overline{A} \neq \emptyset$, $A \cup \overline{A} \neq Z$, $A - \overline{A} \neq \overline{A} - A \neq \emptyset$ (there are also possible cases where $A \subseteq \overline{A}$ and where $\overline{A} \subseteq A$). In addition, the empty multiset \emptyset and the maximal multiset Z are mutually complementary: $\emptyset = \overline{Z}$, $Z = \overline{\emptyset}$, and so $\emptyset \cap \overline{\emptyset} = Z \cap \overline{Z} = \emptyset$.

Another very specific attribute of operations on multisets of family $\mathbf{A} = \{A_i\}_{i \in I}$ is the possibility of their linear combinations with the operation of multiplication by a scalar (reproduction), for example:

weighted union

$$\begin{aligned} \bigcup_{i \in I} (b_i \cdot A_i) &= \{k_{\bigcup_i (b_i \cdot A_i)}(x) \circ x \mid k_{\bigcup_i (b_i \cdot A_i)}(x) \\ &= \max_{i \in I} (b_i \cdot k_{A_i}(x)), b_i \geq 1, \forall x \in X\}; \end{aligned}$$

weighted intersection

$$\begin{aligned} \bigcap_{i \in I} (b_i \cdot A_i) &= \{k_{\bigcap_i (b_i \cdot A_i)}(x) \circ x \mid k_{\bigcap_i (b_i \cdot A_i)}(x) \\ &= \min_{i \in I} (b_i \cdot k_{A_i}(x)), b_i \geq 1, \forall x \in X\}; \end{aligned}$$

weighted arithmetic sum

$$\begin{aligned} \sum_{i \in I} (b_i \cdot A_i) &= \{k_{\sum_i (b_i \cdot A_i)}(x) \circ x \mid k_{\sum_i (b_i \cdot A_i)}(x) \\ &= \sum_{i \in I} (b_i \cdot k_{A_i}(x)), b_i \geq 1, \forall x \in X\}; \end{aligned}$$

weighted arithmetic product

$$\begin{aligned} \prod_{i \in I} (b_i \cdot A_i) &= \{k_{\prod_i (b_i \cdot A_i)}(x) \circ x \mid k_{\prod_i (b_i \cdot A_i)}(x) \\ &= \prod_{i \in I} (b_i \cdot k_{A_i}(x)), b_i \geq 1, \forall x \in X\}; \end{aligned}$$

weighted direct product

$$\begin{aligned} &(b_1 \cdot A_1) \times \dots \times (b_n \cdot A_n) \\ &= \{k_{(b_1 \cdot A_1) \times \dots \times (b_n \cdot A_n)}(x_{p_1}, \dots, x_{p_n}) \circ \langle x_{p_1}, \dots, x_{p_n} \rangle, \\ &\quad b_i \geq 1, \forall x \in X\}; \end{aligned}$$

where the function of the number of instances is given by

$$\begin{aligned} k_{(b_1 \cdot A_1) \times \dots \times (b_n \cdot A_n)}(x_{p_1}, \dots, x_{p_n}) &= \prod_i (b_i \cdot k_{A_i}(x_{p_i})), \\ x_{p_i} &\in A_i, i = 1, \dots, n. \end{aligned}$$

When $b_i = 1$, the “weighted” operations turn into the corresponding “unweighted” operations on multi-

sets. The possibility of introducing and using linear combinations of operations significantly increases the array of means of operating with multisets. When turning from multisets to sets and replacing the multiplicity function $k_A(x)$ by the characteristic function $\chi_A(x)$, many of the above statements for multisets remain valid for sets, but some can become indeterminate, change, or lose meaning.

2. FAMILIES AND FUNCTIONS OF MULTISSETS

Let us consider some special types of families and functions of multisets [6, 7, 11].

The family of all subsets of set A (power set), which includes set A itself and the empty set \emptyset , is called a *Boolean* of set A and is denoted by $P(A)$. Boolean $P(A)$ is a set of all subsets of set A .

A family of all various submultisets of multiset A over domain X (macroset) is called a *macroset* or a *Boolean* of multiset A and is denoted by $\mathbf{P}(A)$. The word *various* plays an important role here, since the theory of multisets allows, generally speaking, unlimited repeatability of elements in the same multiset and, accordingly, the repeatability of multisets in the same family of multisets. In family $\mathbf{P}(A)$, various subsets are always present in a single instance, including multiset A itself, which is the maximal multiset for family $\mathbf{P}(A)$, set $\text{Supp}A$, and the empty multiset \emptyset . Boolean $\mathbf{P}(A)$ is a set of submultisets of multiset A .

Let us call a family of all possible submultisets of multiset A over domain X (power multiset) a *multibool-ean* of multiset A and denote it as $\mathbf{Q}(A)$. Multibool-ean $\mathbf{Q}(A)$, in contrast to Boolean $\mathbf{P}(A)$, is a multiset in which several identical submultisets of multiset A are allowed. Support $\text{Supp} \mathbf{Q}(A)$ of multibool-ean $\mathbf{Q}(A)$ is Boolean $\mathbf{P}(A)$ of multiset A . If $A \subseteq B$, then $\mathbf{P}(A) \subseteq \mathbf{P}(B) \subseteq \mathbf{Q}(A)$. Some of the elements of Booleans $\mathbf{P}(A)$, $\mathbf{P}(B)$, and multibool-ean $\mathbf{Q}(A)$ themselves are sub(multi)sets of other (multi)sets of these families.

The *cardinality of a family* of sets and multisets $P(A)$, $\mathbf{P}(A)$, and $\mathbf{Q}(A)$ is determined by the total number of its elements (respectively, subsets of set A and submultisets of multiset A).

As an example, the cardinality of Boolean $P(A)$, the family of all subsets of the n -element set $A = \{x_1, x_2, \dots, x_n\}$ is equal to $\text{card } P(A) = |P(A)| = 2^n = 2^{|A|} = 2^{\text{card}A}$.

The cardinality of Boolean $\mathbf{P}(A)$, the family of all various subsets of the n -dimensional multiset $A = \{k_A(x_1) \circ x_1, k_A(x_2) \circ x_2, \dots, k_A(x_n) \circ x_n\}$ over a finite n -element set $X = \{x_1, x_2, \dots, x_n\}$ is equal to $\text{card} \mathbf{P}(A) = |\mathbf{P}(A)| = (1 + k_A(x_1)) \cdot (1 + k_A(x_2)) \cdot \dots \cdot (1 + k_A(x_n)) = \prod_{i=1}^n (1 + k_A(x_i))$.

The cardinality of multibool-ean $\mathbf{Q}(A)$, the family of all possible submultisets of an n -dimensional multiset $A = \{k_A(x_1) \circ x_1, k_A(x_2) \circ x_2, \dots, k_A(x_n) \circ x_n\}$ over the

finite n -element set $X = \{x_1, x_2, \dots, x_n\}$, is equal to $\text{card } \mathbf{Q}(A) = |\mathbf{Q}(A)| = 2^m = 2^{|A|} = 2^{\text{card}A}$.

The *capacity* or *volume of a family* of sets and multisets is the sum of cardinalities of (multi)sets that make up this family. Then, respectively, the volume of Boolean $P(A)$ of set A is $\text{vol} P(A) = \|P(A)\| = \sum_{A_j \in P(A)} |A_j| = \sum_{A_j \in P(A)} \sum_x \chi_{A_j}(x)$, the volume of Boolean $\mathbf{P}(A)$ of multiset A over set X is $\text{vol} \mathbf{P}(A) = \|\mathbf{P}(A)\| = \sum_{A_j \in \mathbf{P}(A)} |A_j| = \sum_{A_j \in \mathbf{P}(A)} \left| \sum_x k_{A_j}(x) \right|$, and the volume of multibool-ean $\mathbf{Q}(A)$ of multiset A over set X is $\text{vol} \mathbf{Q}(A) = \|\mathbf{Q}(A)\| = \sum_{A_j \in \mathbf{Q}(A)} |A_j| = \sum_{A_j \in \mathbf{Q}(A)} \sum_x k_{A_j}(x)$.

Thus, Boolean $P(A)$ of set $A = \{a, b, c\}$ consists of sets $\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}$, and $\{a, b, c\}$. The cardinality of Boolean $P(A)$ is equal to $|P(A)| = 2^3 = 8$, the volume $\|P(A)\| = 12$.

Boolean $\mathbf{P}(A)$ of multiset $A = \{3 \circ a, 1 \circ b, 2 \circ c\}$ over set $A = \{a, b, c\}$ includes multisets $\{0 \circ a, 0 \circ b, 0 \circ c\} = \emptyset, \{1 \circ a, 0 \circ b, 0 \circ c\} = \{a\}, \{0 \circ a, 1 \circ b, 0 \circ c\} = \{b\}, \{0 \circ a, 0 \circ b, 1 \circ c\} = \{c\}, \{1 \circ a, 1 \circ b, 0 \circ c\} = \{a, b\}, \{1 \circ a, 0 \circ b, 1 \circ c\} = \{a, c\}, \{0 \circ a, 1 \circ b, 1 \circ c\} = \{b, c\}, \{2 \circ a, 0 \circ b, 0 \circ c\} = \{2 \circ a\}, \{0 \circ a, 0 \circ b, 2 \circ c\}, \{1 \circ a, 1 \circ b, 1 \circ c\} = \{a, b, c\} = \text{Supp}A, \{3 \circ a, 0 \circ b, 0 \circ c\}, \{2 \circ a, 1 \circ b, 0 \circ c\}, \{2 \circ a, 0 \circ b, 1 \circ c\}, \{1 \circ a, 0 \circ b, 2 \circ c\}, \{0 \circ a, 1 \circ b, 2 \circ c\}, \{3 \circ a, 1 \circ b, 0 \circ c\}, \{3 \circ a, 0 \circ b, 1 \circ c\}, \{2 \circ a, 0 \circ b, 2 \circ c\}, \{2 \circ a, 1 \circ b, 1 \circ c\}, \{1 \circ a, 1 \circ b, 2 \circ c\}, \{3 \circ a, 0 \circ b, 2 \circ c\}, \{3 \circ a, 1 \circ b, 1 \circ c\}, \{2 \circ a, 1 \circ b, 2 \circ c\}$, and $\{3 \circ a, 1 \circ b, 2 \circ c\} = A$. The cardinality of Boolean $\mathbf{P}(A)$ is equal to $|\mathbf{P}(A)| = 4 \cdot 2 \cdot 3 = 24$, the volume of $\|\mathbf{P}(A)\| = 72$. In particular, the cardinality of Boolean $P(\emptyset) = \{\emptyset\}$ of the empty (multi)set \emptyset is equal to $|\{\emptyset\}| = 2^0 = 1$, the volume capacity $\|\{\emptyset\}\| = 0$.

Multibool-ean $\mathbf{Q}(A)$ of multiset $A = \{3 \circ a, 1 \circ b, 2 \circ c\}$ over set $A = \{a, b, c\}$ contains 1 multiset $\{0 \circ a, 0 \circ b, 0 \circ c\} = \emptyset$, 3 multisets $\{1 \circ a, 0 \circ b, 0 \circ c\} = \{a\}$, 1 multiset $\{0 \circ a, 1 \circ b, 0 \circ c\} = \{b\}$, 2 multisets $\{0 \circ a, 0 \circ b, 1 \circ c\} = \{c\}$, 3 multisets $\{1 \circ a, 1 \circ b, 0 \circ c\} = \{a, b\}$, 6 multisets $\{1 \circ a, 0 \circ b, 1 \circ c\} = \{a, c\}$, 2 multisets $\{0 \circ a, 1 \circ b, 1 \circ c\} = \{b, c\}$, 3 multisets $\{2 \circ a, 0 \circ b, 0 \circ c\}$, 1 multiset $\{0 \circ a, 0 \circ b, 2 \circ c\}$, 6 multisets $\{1 \circ a, 1 \circ b, 1 \circ c\} = \{a, b, c\} = \text{Supp}A$, 1 multiset $\{3 \circ a, 0 \circ b, 0 \circ c\}$, 3 multisets $\{2 \circ a, 1 \circ b, 0 \circ c\}$, 6 multisets $\{2 \circ a, 0 \circ b, 1 \circ c\}$, 3 multisets $\{1 \circ a, 0 \circ b, 2 \circ c\}$, 1 multiset $\{0 \circ a, 1 \circ b, 2 \circ c\}$, 1 multiset $\{3 \circ a, 1 \circ b, 0 \circ c\}$, 2 multisets $\{3 \circ a, 0 \circ b, 1 \circ c\}$, 3 multisets $\{2 \circ a, 0 \circ b, 2 \circ c\}$, 6 multisets $\{2 \circ a, 1 \circ b, 1 \circ c\}$, 3 multisets $\{1 \circ a, 1 \circ b, 2 \circ c\}$, 1 multiset $\{3 \circ a, 0 \circ b, 2 \circ c\}$, 2 multisets $\{3 \circ a, 1 \circ b, 1 \circ c\}$, 3 multisets $\{2 \circ a, 1 \circ b, 2 \circ c\}$, and 1 multiset $\{3 \circ a, 1 \circ b, 2 \circ c\} = A$. The cardinality of multibool-ean $\mathbf{Q}(A)$ is equal to $|\mathbf{Q}(A)| = 2^6 = 64$, the volume $\|\mathbf{Q}(A)\| = 192$.

For a family of nonempty multisets A_1, \dots, A_p , their union $\bigcup_{s=1}^r A_s = A$ is called *covering* of multiset A , their intersection $\bigcap_{s=1}^r A_s = A$ is *overlapping* of multiset A , and their sum $\sum_{s=1}^r A_s = A$ is *decomposition* of multiset A . If the multisets are pairwise disjoint ($A_p \cap A_q = \emptyset$, $p \neq q$: such multisets are said to be *disjoint*), then covering and decomposition of multiset A coincide and are called *partition* of multiset A . Multisets A_1, \dots, A_r that form the covering, overlapping, decomposition, and partition of multiset A are called its *blocks*, and the number r of blocks is the *rank* of covering, overlapping, decomposition, or partition. Covering of multiset A of rank r is denoted as $\mathbf{C}^r(A) = \{A_1; \dots; A_s; \dots; A_r\}$, overlapping of multiset A of rank r as $\mathbf{I}^r(A) = \{A_1; \dots; A_s; \dots; A_r\}$, decomposition of multiset A of rank r as $\mathbf{D}^r(A) = \{A_1; \dots; A_s; \dots; A_r\}$, and partition of multiset A of rank r as $\mathbf{B}^r(A) = \{A_1; \dots; A_s; \dots; A_r\}$, where blocks are separated by semicolons. If all blocks are submultisets of multiset A , in this case one speaks of partition of multiset A into classes. For the cardinality and volume of covering $\mathbf{C}^r(A)$, overlapping $\mathbf{I}^r(A)$, decomposition $\mathbf{D}^r(A)$, and partition $\mathbf{B}^r(A)$ of multiset A , we have:

$$\begin{aligned} |\mathbf{C}^r(A)| &= |\mathbf{I}^r(A)| = |\mathbf{D}^r(A)| = |\mathbf{B}^r(A)| = r, \\ \|\mathbf{C}^r(A)\| &\geq |A|, \quad \|\mathbf{I}^r(A)\| \geq |A|, \\ \|\mathbf{D}^r(A)\| &= \|\mathbf{B}^r(A)\| = |A|. \end{aligned}$$

Coverings and partitions of a set and coverings, decomposition, and partitions of a multiset can be ordered and unordered, depending on the priority of their blocks.

As an example, a family of zeronos $\{\{0 \circ x_1\}; \dots; \{0 \circ x_s\}; \dots; \{0 \circ x_n\}\}$ is partition $\mathbf{B}^n(\emptyset)$ of an empty (multi)set \emptyset , a family of singletons $\{\{1 \circ x_1\}; \dots; \{1 \circ x_s\}; \dots; \{1 \circ x_n\}\}$ is partition $\mathbf{B}^n(X)$ of the set $X = \{x_1, \dots, x_n\}$, a family of multitons $\{\{k_A(x_1) \circ x_1\}; \dots; \{k_A(x_s) \circ x_s\}; \dots; \{k_A(x_n) \circ x_n\}\}$ is partition $\mathbf{B}^n(A)$ of multiset $A = \{k_A(x_1) \circ x_1, \dots, k_A(x_s) \circ x_s, \dots, k_A(x_n) \circ x_n\}$ with the number of blocks $r = |X| = n$. Boolean $\mathbf{P}(A)$ of set A is covering $\mathbf{C}^r(A)$ of the set A with the number of blocks $r = |\mathbf{P}(A)| = 2^{|A|}$. Boolean $\mathbf{P}(A)$ of the n -dimensional multiset $A = \{k_A(x_1) \circ x_1, \dots, k_A(x_n) \circ x_n\}$ over the finite set $X = \{x_1, \dots, x_n\}$ is covering $\mathbf{C}^r(A)$ of multiset A with the number of blocks $r = |\mathbf{P}(A)| = \prod_{i=1}^n (1 + k_A(x_i))$.

A family of multisets over set $X = \{x_1, x_2, \dots\}$ is called *semiring* \mathbf{E} of multisets if it contains an empty multiset \emptyset , intersection $A_i \cap A_j$ of multisets $A_i, A_j \in \mathbf{E}$, and every multiset $A \in \mathbf{E}$ is representable as a finite union $\bigcup_{s=1}^r A_s = A$ of disjoint submultisets $A_s \subset A$; *ring* \mathbf{K} of

submultisets if it contains an empty multiset \emptyset , union $A_i \cup A_j$, sum $A_i + A_j$, and difference $A_i - A_j$ of multisets $A_i, A_j \in \mathbf{K}$; *algebra* \mathbf{S} of multisets if it is a ring that includes the maximal set, called a unit of algebra, and is closed in relation to finite unions, additions, and complements of multisets; σ -*ring* \mathbf{K}_σ of multisets if it is a ring and contains a countable union $\bigcup_{s=1}^\infty A_s$ and sum $\sum_{s=1}^\infty A_s$ of multisets $A_s \in \mathbf{K}_\sigma$; δ -*ring* \mathbf{K}_δ of multisets if it is a ring and contains a countable intersection $\bigcap_{s=1}^\infty A_s$ of multisets $A_s \in \mathbf{K}_\delta$; σ -*algebra* \mathbf{S}_σ of multisets if it is an σ -ring, contains a unit of algebra, and is closed in relation to countable unions, additions, and complements of multisets.

In particular, every ring of sets/multisets is a semiring of sets/multisets, and every σ -algebra of sets/multisets is a δ -algebra of sets/multisets, and vice versa.

Boolean $\mathbf{P}(A)$ of set A is the minimal algebra of subsets of set A , which is a unit of the algebra. The Boolean algebra $\mathbf{B}(A)$ of subsets of set A is σ -algebra of set X , which is a unit of the algebra, and family $\mathbf{P}(A)$ is the support of the algebra. Boolean $\mathbf{P}(A)$ of the n -dimensional multiset $A = \{k_A(x_1) \circ x_1, \dots, k_A(x_n) \circ x_n\}$ over the finite set $X = \{x_1, \dots, x_n\}$ is the minimal algebra of submultisets of multiset A , which is a unit of algebra. The Boolean algebra $\mathbf{B}(A)$ of submultisets of multiset $A = \{k_A(x_1) \circ x_1, k_A(x_2) \circ x_2, \dots\}$ over the infinite set $X = \{x_1, x_2, \dots\}$ is σ -algebra of multiset A , which is a unit of algebra.

When turning to sets, decomposition of the set will be impossible, and covering, partition, semiring, ring, and algebra of the multiset will turn into the covering, partition, semiring, ring, and algebra of the set, respectively.

A nonnegative real function of multiset $m: \mathbf{A} \rightarrow \mathbb{R}$, defined on family \mathbf{A} , is said to be a *strongly additive* or a *strongly finitely additive measure* if for any finite set of multisets $A_i \in \mathbf{A}$ the following equality holds:

$$m\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n m(A_i), \quad (2)$$

and a *strongly σ -additive* or a *strongly countable additive measure* if for any countable collection of multisets $A_i \in \mathbf{A}$ the following equality holds:

$$m\left(\sum_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty m(A_i). \quad (3)$$

As family \mathbf{A} we have semiring \mathbf{E} , ring \mathbf{K} , σ -ring \mathbf{K}_σ , or σ -algebra \mathbf{S}_σ of submultisets A_i of some multiset $A = \{k_A(x_1) \circ x_1, k_A(x_2) \circ x_2, \dots\}$ over set $X = \{x_1, x_2, \dots\}$. Such a multiset A can be, for example, the maximal multiset $Z = \{k_Z(x_1) \circ x_1, k_Z(x_2) \circ x_2, \dots\}$, constant multiset $Z_{[k]} = \{k \circ x_1, k \circ x_2, \dots\}$, or n -dimensional multiset $X =$

$\{k_X(x_1) \circ x_1, \dots, k_X(x_n) \circ x_n\}$. From the identity property of an empty multiset $\emptyset = \emptyset + \emptyset$ and equality (2) it follows that $m(\emptyset) = m(\emptyset + \emptyset) = 2m(\emptyset)$, and hence $m(\emptyset) = 0$.

For any finite or countable partition of multiset A , the sum of submultisets coincides with their union $\sum_{i=1}^n A_i \cup_{i=1}^n A_i$. Then, expressions (2) and (3) for disjoint multisets $A_i \cap A_j = \emptyset, i \neq j, A_i, A_j \in \mathbf{A}$ can be written as

$$m\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n m(A_i), \tag{4}$$

$$m\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} m(A_i), \tag{5}$$

A multiset measure that satisfies condition (4) will be called *weakly additive* or *weakly finitely additive*, and a measure that satisfies condition (5) will be called *weakly σ -additive* or *weakly countably additive*.

The strongly additive multiset function is also weakly additive. The converse, generally speaking, is not true. The weak additivity (4), (5) of the multiset function coincides with the finite and countable additivity of the set function. For sets, the addition operation is not feasible and there is no strong additivity of the function. Thus, the additivity of multiset functions is more diverse than the additivity of set functions.

The measure m of a multiset also has the properties of monotonicity $m(A) \leq m(B) \Leftrightarrow A \subseteq B$; symmetry $m(A) + m(\bar{A}) = m(Z)$; continuity $\lim_{i \rightarrow \infty} m(A_i) = m(\lim_{i \rightarrow \infty} A_i)$; and elasticity $m(b \cdot A) = bm(A)$, where $b \geq 1$ is an integer. Let the measure of singleton $\{x_i\}$ be equal to $m(\{x_i\}) = w_i, 0 \leq w_i < \infty$. Then, according to equalities (2) and (3), the function defined on the family of multisets A

$$\begin{aligned} m(A) &= m\left(\sum_i A_i\right) = \sum_i m(A_i) \\ &= \sum_i [k_A(x_i) m(\{x_i\})] = \sum_i w_i k_A(x_i) \end{aligned} \tag{6}$$

is strongly additive or, respectively, strongly σ -additive measure of a multiset. If in (6) all the numbers are $w_i = 1$, we obtain: $m(A) = \sum_i k_A(x_i) = |A|$. Thus, the cardinality of a multiset is a special case of its measure.

Let us note the important properties of the multiset measure that are due to its strong additivity and resemble the properties of the cardinality of a multiset:

$$\begin{aligned} m(A + B) &= m(A) + m(B) \\ &= m(A \cup B) + m(A \cap B), \\ m(A - B) &= m(A) - m(A \cap B), \end{aligned} \tag{7}$$

$$\begin{aligned} m(A \Delta B) &= m(A - B) + m(B - A) \\ &= m(A \cup B) - m(A \cap B), \\ m\left(\bigcap_{i=1}^n A_i\right) &\leq m\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n m(A_i). \end{aligned} \tag{8}$$

The *space of multisets with a measure* is defined to be a pair $(\mathbf{S}_\sigma(Z), m)$, where $\mathbf{S}_\sigma(Z)$ is σ -algebra over a maximal multiset Z , m is a σ -finite measure defined on σ -ring \mathbf{K}_σ of submultisets $A_i \subseteq Z$ that form the decomposition of multiset $Z = \sum_{i=1}^{\infty} A_i$. The sum, union, and intersection of a finite number of measurable multisets, the difference and the symmetric difference of two measurable multisets, and the reproduction of a measurable multiset are measurable multisets. Necessary and sufficient conditions for the measurability of a multiset are formulated as follows [6, 7]. Multiset A belonging to σ -ring \mathbf{K}_σ is measurable if and only if for any $\epsilon > 0$ there exists a measurable multiset $B \in \mathbf{K}_\sigma$ such that inequality $m(A \Delta B) < \epsilon$ is satisfied.

3. METRIC SPACES OF MULTISSETS

A set in which the concept of proximity of elements is defined in some way is usually called a space and its elements are called points of space. A nonnegative real function d_X defined on the direct product $X \times X$ of set X is called a *metric* on set X if for any elements the following axioms are satisfied: axioms of (1⁰) symmetry $d_X(x, y) = d_X(y, x)$; (2⁰) identity $d_X(x, y) = 0 \Leftrightarrow x = y$; and (3⁰) triangle $d_X(x, y) \leq d_X(x, z) + d_X(z, y)$. The numerical value of function $d_X(x, y)$ is called the *distance* between elements x and y of set X . Requirements (1⁰)–(3⁰) imply the nonnegativity of metric $d_X(x, y) \geq 0$. Set X with metric d_X defined on it is called a *metric space* and is denoted by (X, d_X) .

A metric space X is said to be *metrically convex* or *d-convex* if for any noncoinciding elements $x, y \in X, x \neq y$ there exists a distinct element $z \in X, z \neq x, z \neq y$ such that the triangle inequality (3⁰) becomes equality $d_X(x, y) = d_X(x, z) + d_X(z, y)$. Metric d_X that satisfies this condition will be called *d-convex*. The property of the metric convexity of the space (X, d_X) is the spatial analog of the ternary relation $[x, z, y]$ “element z lies between elements x and y ,” which in an arbitrary partially ordered set is given by condition $\inf(x, y) \leq z \leq \sup(x, y)$.

The relative positions of the elements of set X are also estimated using other indices of the proximity of the elements, which are based on the incomplete axiomatics of the metric space. The nonnegative real function d_X defined on the direct product $X \times X$, which satisfies the symmetry axiom (1⁰) and the coincidence condition (4⁰) $d_X(x, x) = 0$ for all $x \in X$, is called a *quasi-metric*, and the one that satisfies the triangle axiom

(3⁰) is called a *pseudometric* on set X . The corresponding spaces (X, d_X) are called *quasimetric* and *pseudometric*. In the quasi- and pseudometric spaces the identity axiom (2⁰) for function d_X is not satisfied, that is, condition $d_X(x, y) = 0$ does not, in general, imply the equality of elements x and y . The coincidence condition (4⁰) is weaker than the identity axiom (2⁰). Therefore, every metric is also a quasimetric and a pseudometric. The converse is not true in the general case.

A nonnegative real function d_X defined on the direct product $X \times X$ and that satisfies the symmetry (1⁰) and identity (2⁰) axioms is called a *symmetric* on set X , and set X with the symmetric d_X is called a *proximity space*. Symmetric d_X that satisfies the inequality (5⁰) $d_X(x, y) \leq \max_{z \in X} [d_X(x, z), d_X(z, y)]$ for any elements of set X is called the *ultrametric* on set X or the ultrametric distance between elements x and y . It is easy to verify that the triangle axiom (3⁰) also holds for the ultrametric. Thus, ultrametric d_X satisfies the axiomatics of the metric space, and space (X, d_X) with ultrametric d_X is metric. The inequality (5⁰) is stronger than the triangle inequality (3⁰). Therefore, any ultrametric is also a metric. The converse is not true in the general case.

Let us consider possible approaches to the formation of metric spaces (A, d_A) on families of A sets and A multisets. Any set $A \subseteq X$ and multiset $A \subseteq Z$ over set $X = \{x_1, \dots, x_n\}$ can be associated with an integer vector $\mathbf{y} = (y_1, \dots, y_n)$ whose components are given by, respectively, as $y_i = \chi_A(x_i) = 0, 1$ for set A and $y_i = k_A(x_i) = 0, 1, 2, \dots$ for multiset $A, x_i \in X$. As an example, vector (1,1,1,1,1,1,1) corresponds to set $X = \{a, b, c, d, e, f, g\}$, vector (0,0,0,0,0,0,0) corresponds to the empty (multi)set \emptyset , vector (1,0,0,0,0,0,0) corresponds to singleton $\{a\}$, vector (0,2,0,0,0,0,0) corresponds to multiton $\{2 \circ b\}$, vector (1,1,1,0,0,0,0) corresponds to set $A = \{a, b, c\}$, and vector (3,1,2,0,0,0,0) corresponds to multiset $A = \{3 \circ a, 1 \circ b, 2 \circ c\}$.

We use an analogy with known types of metric spaces: vector spaces l_p^n and spaces of bounded numerical sequences l_p [1].

Boolean $\mathbf{P}(X)$ of various submultisets of an n -dimensional multiset $X = \{k_X(x_1) \circ x_1, \dots, k_X(x_n) \circ x_n\}$ over the finite set $X = \{x_1, \dots, x_n\}$ is equivalent to set \mathbb{R}^n of n -dimensional vectors $\mathbf{y}_s = (y_{s_1}, \dots, y_{s_n})$ with real components $y_{s_i} = k_A(x_i)$ and forms a *metric space of finite multisets* [19]:

$\mathbf{P}_1 = (\mathbf{P}(X), d_{P_1})$ with the Hamming-type metric

$$d_{P_1}(A, B) = \sum_{i=1}^n |k_A(x_i) - k_B(x_i)|; \quad (9)$$

$\mathbf{P}_2 = (\mathbf{P}(X), d_{P_2})$ with the Euclidean-type metric

$$d_{P_2}(A, B) = \left[\sum_{i=1}^n |k_A(x_i) - k_B(x_i)|^2 \right]^{1/2}; \quad (10)$$

$\mathbf{P}_p = (\mathbf{P}(X), d_{P_p})$ with the Minkowski-type metric

$$d_{P_p}(A, B) = \left[\sum_{i=1}^n |k_A(x_i) - k_B(x_i)|^p \right]^{1/p}; \quad (11)$$

$\mathbf{P}_\infty = (\mathbf{P}(X), d_{P_\infty})$ with the Chebyshev-type metric

$$d_{P_\infty}(A, B) = \max_i |k_A(x_i) - k_B(x_i)|. \quad (12)$$

Here, $p \geq 1$ is an integer.

σ -algebra $\mathbf{S}_\sigma(X)$ of different submultisets of multiset $X = \{k_X(x_1) \circ x_1, k_X(x_2) \circ x_2, \dots\}$, over an arbitrary set $X = \{x_1, x_2, \dots\}$, where any submultiset $A_s \subseteq X$ satisfies the boundedness condition $k_{A_s}(x_i) \leq c_s < \infty$ or $\sum_{i=1}^\infty (k_{A_s}(x_i))^p \leq c^p < \infty$ for all integers $p \geq 1$, is equivalent to set \mathbb{R}^N of bounded numerical sequences $Y_s = \{y_{s_i}\} = \{y_{s_1}, y_{s_2}, \dots\}$ with real terms $y_{s_i} = k_{A_s}(x_i)$ that satisfies the condition $|y_{s_i}| \leq c_s < \infty$ or $\sum_{i=1}^\infty |y_{s_i}|^p \leq c^p < \infty$ for all integers $p \geq 1$ and forms a *metric space of bounded multisets*:

$\mathbf{S}_1 = (\mathbf{S}_\sigma(X), d_{S_1})$ with the Hamming-type metric

$$d_{S_1}(A, B) = \sum_{i=1}^\infty |k_A(x_i) - k_B(x_i)|; \quad (13)$$

$\mathbf{S}_2 = (\mathbf{S}_\sigma(X), d_{S_2})$ with the Euclidean-type metric

$$d_{S_2}(A, B) = \left[\sum_{i=1}^\infty |k_A(x_i) - k_B(x_i)|^2 \right]^{1/2}; \quad (14)$$

$\mathbf{S}_p = (\mathbf{S}_\sigma(X), d_{S_p})$ with the Minkowski-type metric

$$d_{S_p}(A, B) = \left[\sum_{i=1}^\infty |k_A(x_i) - k_B(x_i)|^p \right]^{1/p}; \quad (15)$$

$\mathbf{S}_\infty = (\mathbf{S}_\sigma(X), d_{S_\infty})$ with the Chebyshev-type metric

$$d_{S_\infty}(A, B) = \sup_i |k_A(x_i) - k_B(x_i)|. \quad (16)$$

Let us note that expressions (9)–(11) and (13)–(15) comprise the cardinality $|\mathbf{A} \Delta \mathbf{B}| = \sum_i |k_A(x_i) - k_B(x_i)|$ of the symmetric difference of multisets A and B , which plays the role of the metric. Replacing the multiplicity function $k_A(x)$ by the characteristic function $\chi_A(x)$ in formulas (9)–(16) we obtain definitions of metric spaces of finite and bounded sets.

The space $(\mathbf{S}_\sigma(Z), m)$ of multisets with a measure, where multisets $A_s \subseteq Z$ that comprise the σ -algebra $\mathbf{S}_\sigma(Z)$ of the maximal multiset $Z = \{k_Z(x_1) \circ x_1, k_Z(x_2) \circ x_2, \dots\}$ over set $X = \{x_1, x_2, \dots\}$ are measurable and measure m of the multiset is strongly σ -additive and com-

pletely σ -finite, forms the *metric spaces of measurable multisets* $\mathbf{Z}_{qp} = (\mathbf{S}_\sigma(\mathbf{Z}), d_{\mathbf{Z}_{qp}})$, where $q = 1, 2, 3, 4, p \geq 1$ is an integer, with the functions

$$d_{\mathbf{Z}_{1p}}(\mathbf{A}, \mathbf{B}) = [m(\mathbf{A}\Delta\mathbf{B})]^{1/p}, \quad (17)$$

$$d_{\mathbf{Z}_{2p}}(\mathbf{A}, \mathbf{B}) = [m(\mathbf{A}\Delta\mathbf{B})/m(\mathbf{Z})]^{1/p}, \quad (18)$$

$$d_{\mathbf{Z}_{3p}}(\mathbf{A}, \mathbf{B}) = [m(\mathbf{A}\Delta\mathbf{B})/m(\mathbf{A} \cup \mathbf{B})]^{1/p}, \quad (19)$$

$$d_{\mathbf{Z}_{4p}}(\mathbf{A}, \mathbf{B}) = [m(\mathbf{A}\Delta\mathbf{B})/m(\mathbf{A} + \mathbf{B})]^{1/p}. \quad (20)$$

Functions $d_{\mathbf{Z}_{3p}}$ and $d_{\mathbf{Z}_{4p}}$ are not defined for $\mathbf{A} = \mathbf{B} = \emptyset$; therefore, by definition, we take $d_{\mathbf{Z}_{3p}}(\emptyset, \emptyset) = d_{\mathbf{Z}_{4p}}(\emptyset, \emptyset) = 0$. Function $d_{\mathbf{Z}_{1p}}$ defines mapping $d_{\mathbf{Z}_{1p}}: \mathbf{S}_\sigma \times \mathbf{S}_\sigma \rightarrow \mathbb{R}_+$, and functions $d_{\mathbf{Z}_{2p}}, d_{\mathbf{Z}_{3p}}$, and $d_{\mathbf{Z}_{4p}}$ define mapping $d_{\mathbf{Z}_{qp}}: \mathbf{S}_\sigma \times \mathbf{S}_\sigma \rightarrow \mathbb{R}_{01} = [0, 1]$. Functions $d_{\mathbf{Z}_{qp}}(\mathbf{A}, \mathbf{B}), q = 1, 2, 3$ are pseudometrics that satisfy the symmetry (1⁰) and triangle (3⁰) axioms and the coincidence condition (4⁰), and function $d_{\mathbf{Z}_{4p}}(\mathbf{A}, \mathbf{B})$ is a quasimetric that satisfies the symmetry axiom (1⁰) and the coincidence condition (4⁰). In the general case, condition $m(\mathbf{A}\Delta\mathbf{B}) = 0$ does not imply that $\mathbf{A} = \mathbf{B}$. For multisets that differ by a multiset of measure zero, condition $m(\mathbf{A}\Delta\mathbf{B}) = 0$ implies the m -equality of such multisets $\mathbf{A} = {}_m\mathbf{B}$, and then for functions $d_{\mathbf{Z}_{qp}}(\mathbf{A}, \mathbf{B}), q = 1, 2, 3, 4$, the identity axiom (2⁰) also holds. Functions $d_{\mathbf{Z}_{1p}}, d_{\mathbf{Z}_{2p}}$, and $d_{\mathbf{Z}_{3p}}$ become metrics, and function $d_{\mathbf{Z}_{4p}}$ is symmetric almost everywhere on σ -algebra $\mathbf{S}_\sigma(\mathbf{Z})$ of measurable multisets. The proofs of these statements for pseudometrics were given in [6, 7], and for a quasimetric, in the appendix.

Let us call pseudometric $d_{\mathbf{Z}_{1p}}(\mathbf{A}, \mathbf{B})$ on space $(\mathbf{S}_\sigma(\mathbf{Z}), m)$ of measurable multisets *general*, pseudometric $d_{\mathbf{Z}_{2p}}(\mathbf{A}, \mathbf{B})$ *completely averaged*, pseudometric $d_{\mathbf{Z}_{3p}}(\mathbf{A}, \mathbf{B})$ *locally averaged*, and quasimetric $d_{\mathbf{Z}_{4p}}(\mathbf{A}, \mathbf{B})$ *averaged*. The completely averaged pseudometric $d_{\mathbf{Z}_{2p}}(\mathbf{A}, \mathbf{B})$ characterizes the proximity of two multisets \mathbf{A} and \mathbf{B} , referred to the maximal possible distance in the original space. The locally averaged pseudometric $d_{\mathbf{Z}_{3p}}(\mathbf{A}, \mathbf{B})$ characterizes the proximity of two multisets \mathbf{A} and \mathbf{B} , referred to the joint “common part” of these two multisets in the original space. The averaged quasimetric $d_{\mathbf{Z}_{4p}}(\mathbf{A}, \mathbf{B})$ characterizes the proximity of two multisets \mathbf{A} and \mathbf{B} , referred to the maximal possible “common part” of these two multisets in the original space.

The basic pseudometric $d_{\mathbf{Z}_{1p}}(\mathbf{A}, \mathbf{B}) = [m(\mathbf{A}\Delta\mathbf{B})]^{1/p}$ and the completely averaged pseudometric $d_{\mathbf{Z}_{2p}}(\mathbf{A}, \mathbf{B}) = [m(\mathbf{A}\Delta\mathbf{B})/m(\mathbf{Z})]^{1/p}$ are continuous, uniformly continuous, and equicontinuous functions, and the locally averaged

pseudometric $d_{\mathbf{Z}_{3p}}(\mathbf{A}, \mathbf{B}) = [m(\mathbf{A}\Delta\mathbf{B})/m(\mathbf{A} \cup \mathbf{B})]^{1/p}$ and the averaged quasimetric $d_{\mathbf{Z}_{4p}}(\mathbf{A}, \mathbf{B}) = [m(\mathbf{A}\Delta\mathbf{B})/m(\mathbf{A} + \mathbf{B})]^{1/p}$ are piecewise continuous functions of their variables almost everywhere on the corresponding metric space for any number p [6, 7]. Pseudometrics $d_{\mathbf{Z}_{qp}}(\mathbf{A}, \mathbf{B}), q = 1, 2, 3$ and quasimetrics $d_{\mathbf{Z}_{4p}}(\mathbf{A}, \mathbf{B})$ are not similar to any metrics of other metric spaces, in particular, to Minkowski-type metrics (11), (15). When $w_i = 1$ for all i in formula (6), the general pseudometric $d_{\mathbf{Z}_{11}}(\mathbf{A}, \mathbf{B}) = |\mathbf{A}\Delta\mathbf{B}|$ coincides with the Hamming-type metrics (9), (13).

The metric spaces of multisets have the following properties [6, 7]. The spaces $\mathbf{P}_p, \mathbf{P}_\infty$ of finite multisets are metrically convex, l_p^n - and l_p -embeddable, everywhere dense, separable, incomplete, noncompact, but relatively compact. The spaces $\mathbf{S}_p, \mathbf{S}_\infty$ of bounded multisets are metrically convex, separable, complete, noncompact, but locally compact. The spaces $\mathbf{Z}_{1p} = (\mathbf{S}_\sigma(\mathbf{Z}), d_{\mathbf{Z}_{1p}}), \mathbf{Z}_{2p} = (\mathbf{S}_\sigma(\mathbf{Z}), d_{\mathbf{Z}_{2p}})$ of measurable multisets are homeomorphic, complete, and separable. The spaces $\mathbf{Z}_{11} = (\mathbf{S}_\sigma(\mathbf{Z}), d_{\mathbf{Z}_{11}}), \mathbf{Z}_{21} = (\mathbf{S}_\sigma(\mathbf{Z}), d_{\mathbf{Z}_{21}}), \mathbf{Z}_{31} = (\mathbf{S}_\sigma(\mathbf{Z}), d_{\mathbf{Z}_{31}})$ of measurable multisets are metrically convex, and, in addition, the spaces \mathbf{Z}_{11} and \mathbf{Z}_{21} are isometric to the space of bounded numerical sequences l_1 .

When the multiplicity function $k_A(x)$ is replaced by the characteristic function $\chi_A(x)$, the metric spaces of measurable multisets are transformed into the metric spaces of measurable sets that have many similar properties. The general metric $d_{X_{11}}(\mathbf{A}, \mathbf{B}) = m(\mathbf{A}\Delta\mathbf{B})$ or $d_{X_{31}}(\mathbf{A}, \mathbf{B}) = |\mathbf{A}\Delta\mathbf{B}|$ is called the Fréchet–Nikodym–Aronshayan distance or a metric of measure. The locally averaged metric $d_{X_{31}}(\mathbf{A}, \mathbf{B}) = m(\mathbf{A}\Delta\mathbf{B})/m(\mathbf{A} \cup \mathbf{B})$ is called the Steinhaus distance and $d_{X_{31}}(\mathbf{A}, \mathbf{B}) = |\mathbf{A}\Delta\mathbf{B}|/|\mathbf{A} \cup \mathbf{B}|$ is called the biotopic distance [1–3]. The spaces of measurable sets and multisets with metrics of the form (17)–(20) were first introduced in our papers [8–10].

4. REPRESENTATION AND COMPARISON OF MULTI-ATTRIBUTE OBJECTS

Let us discuss possible ways of representing and comparing objects that are present in several differing instances (versions or copies) and are described by many quantitative and qualitative attributes [12, 15].

Let us first consider the situation where the available objects A_1, \dots, A_m are in single instances and are characterized by attributes K_1, \dots, K_n with numerical and/or verbal scales of estimates $X_l = \{x_l^1, \dots, x_l^{h_l}\}, l = 1, \dots, n$. Traditionally, each object $A_i, i = 1, \dots, m$ is associated with a vector or tuple $x_i = (x_{i_1}^{e_1}, \dots, x_{i_n}^{e_n})$,

where $x_i^{e_l}$ is one of the gradations on scale X_l of attribute K_l . Vector/tuple \mathbf{x} is a point of the n -dimensional space $X = X_1 \times \dots \times X_n$ formed by the scales of estimates of attributes K_1, \dots, K_n . The collection of multi-attribute objects A_1, \dots, A_m and their attributes can also be represented by the table Objects–Attributes $F = \|x_i\|_{m \times n}$. Rows of matrix F correspond to objects, columns correspond to attributes, and elements x_i of the matrix are values of components $x_i^{e_l}$ of vectors/tuples that specify objects.

The situation becomes significantly more complicated when the same multi-attribute object A_i , $i = 1, \dots, m$ is present in several instances (versions or copies) $A_i^{(s)}$, $s = 1, \dots, t$, which differ in the attribute values K_1, \dots, K_n . In such cases, the object A_i will be associated not with one vector/tuple but a group of t vectors/tuples $\{\mathbf{x}_i^{(1)}, \dots, \mathbf{x}_i^{(t)}\}$. Here, the vector/tuple $\mathbf{x}_i^{(s)} = (x_i^{(s)}, \dots, x_n^{(s)})$ describes one of the versions $A_i^{(s)}$ of object A_i , and the component $x_i^{(s)} = K_l(A_i^{(s)})$ is the attribute value K_l in the s -th version $A_i^{(s)}$ of object A_i , which is equal to x^j , $j = 1, \dots, h$ if all attributes K_1, \dots, K_n have the same scale of estimates $X = \{x^1, \dots, x^h\}$, or $x_i^{e_l}$, $e_l = 1, \dots, h_l$ if attributes K_1, \dots, K_n have their own scales of estimates $X_l = \{x_l^1, \dots, x_l^{h_l}\}$, $l = 1, \dots, n$.

The collection of objects A_1, \dots, A_m , each of which exists in several versions $A_i^{(s)}$, and their attributes can again be represented by the table Objects–Attributes $F^\# = \|x_i^{(s)}\|_{tm \times n}$, but have a larger dimension. The number of rows of matrix $F^\#$ is equal to the number of all versions of objects, the number of columns is equal to the number of attributes, and elements $x_i^{(s)}$ of the matrix are values of components $x_i^{(s)}$ of vectors/tuples that specify different versions of objects.

Object A_i is now represented in the n -dimensional space of attributes $X = X_1 \times \dots \times X_n$ not by a single point \mathbf{x}_i but by an entire group (cloud) consisting of t points $\{\mathbf{x}_i^{(1)}, \dots, \mathbf{x}_i^{(t)}\}$. It is important to note that the group of vectors/tuples $\mathbf{x}_i^{(1)}, \dots, \mathbf{x}_i^{(t)}$ that represent object A_i should be treated as a single whole. In this case, generally speaking, individual values of attributes of different versions of object A_i (estimates of different experts, characteristics measured in different ways or instruments) can be similar, different, and even contradictory, which in turn can lead to incomparability of vectors/tuples that represent the same object A_i .

The collection of such multi-attribute objects A_1, \dots, A_m , each of which is represented in the attribute space $X = X_1 \times \dots \times X_n$ by its cloud consisting of t different

points, can have a rather complex structure that is difficult to analyze. Therefore, it is highly desirable to somehow simplify the description and aggregate the representation of such multi-attribute objects.

One possible approach to simplifying the description of the multi-attribute object A_i represented by a group of vectors $\mathbf{x}_i^{(1)}, \dots, \mathbf{x}_i^{(t)}$ is as follows. Since components of any vector $x_i^{(s)} = K_l(A_i^{(s)})$ are numerical values, in the attribute space $X = X_1 \times \dots \times X_n$, instead of a group of vectors, one can specify a single vector $\mathbf{x}_i^{\text{cond}} = (x_i^{\text{cond}}, \dots, x_n^{\text{cond}})$, which will correspond to object A_i .

The components of vector $\mathbf{x}_i^{\text{cond}}$ are determined from some additional significant considerations or formal conditions. As an example, it can be a vector that is the center of a group, a vector closest to all vectors in a group, or a vector with averaged or weighted values of components of vectors in a group.

However, when the attribute scales have discrete numerical estimates, the “averaged” vector may not physically exist in the original n -dimensional attribute space $X = X_1 \times \dots \times X_n$, since there are no corresponding numerical grades on the scales. In order to be able to operate with such vectors, it is necessary either to expand the initial scales of estimates by introducing intermediate numerical grades or to consider the scales of estimates to be continuous. Both, strictly speaking, change the original formulation of the problem. When objects are represented by tuples with symbolic, verbal or mixed values of components, a group of such tuples in principle cannot be replaced by a single tuple with averaged, weighted, or mixed estimates since such operations are mathematically not feasible.

If the group consists of only two vectors/tuples, then it is not possible to introduce some single object as a typical representative of the group, for example, one closest to all objects of the group. This is true for both numeric and symbolic attribute scales.

These and other similar difficulties can be overcome if we use the formalism of the theory of multisets, which allows us to take simultaneously heterogeneous attributes, possible combinations of attribute values, and the presence of different versions of objects into account.

Let all attributes K_1, \dots, K_n have the same scale of estimates $X = \{x^1, \dots, x^h\}$. Let us choose set X as a generating set and associate each object A_i , $i = 1, \dots, m$ with a multiset of the form (1)

$$A_i = \{k_{A_i}(x^1) \circ x^1, \dots, k_{A_i}(x^h) \circ x^h\} \quad (21)$$

over set $X = \{x^1, \dots, x^h\}$. Here, the value of the multiplicity function $k_{A_i}(x^j)$ shows how many times estimate x^j , $j = 1, \dots, h$ is present in the description of object A_i . In this case, the collection of objects A_1, \dots, A_m and their attributes can be represented by the table

Objects—Attributes $G = \|k_{ij}\|_{m \times h}$. The rows of matrix G correspond to objects, the columns correspond to values of the X scale of attributes, and the elements k_{ij} correspond to the multiplicity values $k_{A_i}(x^j)$ of x^j elements of the multisets defining the objects.

Let us suppose now that each attribute K_l has its own scale of estimates $X_l = \{x_l^1, \dots, x_l^{h_l}\}$, $l = 1, \dots, n$. Let us introduce a single scale (hyperscale) of attributes: set $X = X_1 \cup \dots \cup X_n = \{x_1^1, \dots, x_1^{h_1}, \dots; x_n^1, \dots, x_n^{h_n}\}$, which consists of n groups of attributes and combines all gradations of the estimates on the scales of all attributes. Then, each object A_i will be associated with the multiset

$$A_i = \{k_{A_i}(x_1^1) \circ x_1^1, \dots, k_{A_i}(x_1^{h_1}) \circ x_1^{h_1}, \dots; k_{A_i}(x_n^1) \circ x_n^1, \dots, k_{A_i}(x_n^{h_n}) \circ x_n^{h_n}\} \quad (22)$$

over set $X = \{x_1^1, \dots, x_1^{h_1}, \dots; x_n^1, \dots, x_n^{h_n}\}$ of gradations of scales of criteria. Here, the value of the multiplicity function $k_{A_i}(x_l^{e_l})$ shows how many times estimate $x_l^{e_l} \in X_l$, $e_l = 1, \dots, h_l$ for attribute K_l is present in the description of object A_i .

Despite the seemingly cumbersome representation of multi-attribute objects using multisets, these forms of notation are extremely convenient when comparing objects and performing operations, since calculations are carried out in parallel and simultaneously on all elements of all multisets. Let us note that expression (22) can also be easily written in the “usual” form (21) if we make a change of variables in set $X = \{x_1^1, \dots, x_1^{h_1}, \dots; x_n^1, \dots, x_n^{h_n}\}$: $x_1^1 = x^1, \dots, x_1^{h_1} = x^{h_1}, x_2^1 = x^{h_1+1}, \dots, x_2^{h_2} = x^{h_1+h_2}, \dots, x_n^{h_n} = x^h$, $h = h_1 + \dots + h_n$.

The collection of objects A_1, \dots, A_m and values of their attributes, represented by multisets of the form (22), can be defined by another table, Object—Attributes $H = \|k_{ij}\|_{m \times h}$, $h = h_1 + \dots + h_n$. The rows of matrix H correspond to objects, the columns correspond to values of the attribute scales X_j , and the elements k_{ij} correspond to the multiplicity values $k_{A_i}(x_l^{e_l})$ of elements $x_l^{e_l}$ of the multisets defining the objects.

The variety of operations over multisets provides the ability to group multi-attribute objects in different ways. It is common to combine objects into groups using the operation of adding vectors or combining sets. In the case of multisets, group D_f of objects can be formed as a sum $C_f = \sum_i A_i$, $k_{C_f}(x^j) = \sum_i k_{A_i}(x^j)$, union $C_f = \bigcup_i A_i$, $k_{C_f}(x^j) = \max_i k_{A_i}(x^j)$, intersection $C_f = \bigcap_i A_i$, $k_{C_f}(x^j) = \min_i k_{A_i}(x^j)$, or as a linear combination $C_f = \sum_i b_i \cdot A_i$, $C_f = \bigcup_i b_i \cdot A_i$, $C_f = \bigcap_i b_i \cdot A_i$

of multisets that define the objects. When group D_f is formed by adding multisets, all properties (all values of all attributes) of all terms of the group are aggregated. When multisets are united or intersected, the best properties (maximum values of all attributes) or, respectively, the worst properties (minimum values of all attributes) for individual terms of the group increase.

Let us associate each version $A_i^{(s)}$, $i = 1, \dots, m$, $s = 1, \dots, t$ of the multi-attribute object A_i with multiset $A_i^{(s)} = \{k_{A_i}^{(s)}(x^1) \circ x^1, \dots, k_{A_i}^{(s)}(x^h) \circ x^h\}$, and each object A_i with multiset $A_i = \{k_{A_i}(x^1) \circ x^1, \dots, k_{A_i}(x^h) \circ x^h\}$ over the set of estimates $X = \{x^1, \dots, x^h\}$ or $X = X_1 \cup \dots \cup X_n = \{x_1^1, \dots, x_1^{h_1}, \dots; x_n^1, \dots, x_n^{h_n}\}$. Multiset A_i representing object A_i as a whole will be formed by weighted addition of the multisets that describe the versions of this object: $A_i = c^{(1)} A_i^{(1)} + \dots + c^{(t)} A_i^{(t)}$, where the multiplicity function is calculated by the rule $k_{A_i}(x^j) = \sum_s c^{(s)} k_{A_i}^{(s)}(x^j)$, $j = 1, \dots, h$. $c^{(s)}$ characterizes the significance of the s -th version (expert competence or measurement accuracy) whose values are sometimes normalized by $\sum_s c^{(s)} = 1$.

The objects A_1, \dots, A_m , each of which exists in several versions, and the values of whose attributes are represented by multisets of the form (21) or (22), can be specified by the tables Objects—Attributes $G^\# = \|k_{ij}\|_{m \times h}$, $H^\# = \|k_{ij}\|_{m \times h}$, or $H = \|k_{ij}\|_{m \times h}$, $h = h_1 + \dots + h_n$. Elements k_{ij} of matrices $G^\#, H^\#$, or H are multiplicities $k_{A_i}^{(s)}(x^j)$, $k_{A_i}^{(s)}(x_l^{e_l})$, or $k_{A_i}(x_l^{e_l})$ of elements of the multisets that describe the corresponding versions $A_i^{(s)}$ of objects A_i or the objects A_i themselves.

5. AN ILLUSTRATIVE EXAMPLE

There is a collection consisting of ten objects A_1, \dots, A_{10} described by eight attributes K_1, \dots, K_8 , each of which takes one of the values on a five-point scale of estimates $X = \{x^1, x^2, x^3, x^4, x^5\}$ [15]. As an example, objects A_1, \dots, A_{10} are survey questions for studying public opinion on a certain issue. Then, the attributes of the objects are the answers of eight respondents K_1, \dots, K_8 , which are coded as follows: x^1 , 1/totally disagree; x^2 , 2/disagree; x^3 , 3/neutral; x^4 , 4/agree; x^5 , 5/completely agree. Alternately, the objects A_1, \dots, A_{10} could be schoolchildren and the object attributes are their annual scores in eight subjects: K_1 , Mathematics; K_2 , Physics; K_3 , Chemistry; K_4 , Biology; K_5 , Geography; K_6 , History; K_7 , Literature; K_8 , Foreign Language. The gradations of the estimates mean the following: x^1 , 1/very bad; x^2 , 2/bad; x^3 , 3/satisfactory; x^4 , 4/good; x^5 , 5/excellent.

Table 1, Objects–Attributes F, borrowed from [5] characterizes answers of university students who evaluated a course of lectures using numerical estimates. The same results of a student survey or schoolchildren scores recorded as multisets of numerical or verbal estimates are presented in Table 2, Objects–Attributes G. As an example, in Table 1 the answers of eight students to question A_1 are given by the vector/tuple of estimates $x_1 = (4, 5, 4, 5, 4, 5, 4, 5)$, while in Table 2 they are given by a multiset of estimates of the form (21): $A_1 = \{0 \circ x^1, 0 \circ x^2, 0 \circ x^3, 4 \circ x^4, 4 \circ x^5\}$. The last form of notation shows that object A_1 has four estimates x^4 that mean “agree” or “good,” four estimates x^5 that mean “completely agree” or “excellent,” and there are no other estimates for object A_1 . If object A_1 is characterized by estimates for each of attributes K_1, \dots, K_8 with different scales, then A_1 is represented by a multiset of the form (22):

$$A_1 = \{0 \circ x_1^1, 0 \circ x_1^2, 0 \circ x_1^3, 1 \circ x_1^4, 0 \circ x_1^5; \\ 0 \circ x_2^1, 0 \circ x_2^2, 0 \circ x_2^3, 0 \circ x_2^4, 1 \circ x_2^5; \\ 0 \circ x_3^1, 0 \circ x_3^2, 0 \circ x_3^3, 1 \circ x_3^4, 0 \circ x_3^5; \\ 0 \circ x_4^1, 0 \circ x_4^2, 0 \circ x_4^3, 0 \circ x_4^4, 1 \circ x_4^5; \\ 0 \circ x_5^1, 0 \circ x_5^2, 0 \circ x_5^3, 1 \circ x_5^4, 0 \circ x_5^5; \\ 0 \circ x_6^1, 0 \circ x_6^2, 0 \circ x_6^3, 0 \circ x_6^4, 1 \circ x_6^5; \\ 0 \circ x_7^1, 0 \circ x_7^2, 0 \circ x_7^3, 1 \circ x_7^4, 0 \circ x_7^5; \\ 0 \circ x_8^1, 0 \circ x_8^2, 0 \circ x_8^3, 0 \circ x_8^4, 1 \circ x_8^5\}.$$

Now, let each object be present in several instances (versions or copies) differing between each other. As an example, eight respondents K_1, \dots, K_8 , who gave answers to the same questions A_1, \dots, A_{10} , were polled twice. Alternatively, schoolchildren A_1, \dots, A_{10} received scores for the same eight disciplines K_1, \dots, K_8 in each semester, that is, twice a year. Thus, each object will be specified not by one but by two vectors/tuples of attributes, or by two multisets. Formally, we can assume that the method used to describe versions of the object is an expression of the individual opinion of an expert, and the description of the object “in general” is an aggregate collective judgment of two experts.

The different versions of the objects are defined in Tables 3 and 4, Objects–Attributes $F^\#$ and $G^\#$, respectively. In Table 3, object A_1 is represented by a group of two vectors/tuples $x_1^{(1)} = (4, 5, 4, 5, 4, 5, 4, 5)$ and $x_1^{(2)} = (5, 5, 5, 5, 4, 4, 4, 5)$. The aggregated estimates of object A_1 , when averaged, will be described by vector $x_1^{av} = (4.5, 5.0, 4.5, 5.0, 4.0, 4.5, 4.0, 5.0)$ of the “average” estimates, although non-integer numbers are absent in the established five-point scale $X = \{1, 2, 3, 4, 5\}$. Two versions of object A_1 are represented in Table 4 by multisets of estimates of the form (21) $A_1^{(1)} =$

Table 1. Data table F

$A \setminus K$	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8
A_1	4	5	4	5	4	5	4	5
A_2	4	1	2	1	3	2	2	2
A_3	1	1	3	1	4	1	1	4
A_4	5	3	2	4	4	5	4	5
A_5	4	4	4	4	4	5	4	4
A_6	5	5	4	4	4	5	5	4
A_7	4	1	2	3	3	3	1	2
A_8	4	5	4	2	3	4	5	3
A_9	3	2	3	1	3	3	2	2
A_{10}	5	5	4	5	3	5	5	4

Table 2. Data table G

$A \setminus X$	x^1	x^2	x^3	x^4	x^5
A_1	0	0	0	4	4
A_2	2	4	1	1	0
A_3	5	0	1	2	0
A_4	0	1	1	3	3
A_5	0	0	0	7	1
A_6	0	0	0	4	4
A_7	2	2	3	1	0
A_8	0	1	2	3	2
A_9	1	3	4	0	0
A_{10}	0	0	1	2	5

$\{0 \circ x^1, 0 \circ x^2, 0 \circ x^3, 4 \circ x^4, 4 \circ x^5\}$ and $A_1^{(2)} = \{0 \circ x^1, 0 \circ x^2, 0 \circ x^3, 3 \circ x^4, 5 \circ x^5\}$. For simplicity, we will consider the respondents as equally competent or semester scores to be equally significant: $c^{(1)} = c^{(2)} = 1$. Then, object A_1 “as a whole” is described by multiset

$$A_1 = A_1^{(1)} + A_1^{(2)} = \{0 \circ x^1, 0 \circ x^2, 0 \circ x^3, 7 \circ x^4, 9 \circ x^5\}.$$

Similarly, objects $A_2, A_4, A_6, A_7,$ and A_8 are defined by multisets

$$A_2 = A_2^{(1)} + A_2^{(2)} = \{4 \circ x^1, 6 \circ x^2, 4 \circ x^3, 2 \circ x^4, 0 \circ x^5\},$$

$$A_4 = A_4^{(1)} + A_4^{(2)} = \{0 \circ x^1, 1 \circ x^2, 3 \circ x^3, 7 \circ x^4, 5 \circ x^5\},$$

$$A_6 = A_6^{(1)} + A_6^{(2)} = \{0 \circ x^1, 0 \circ x^2, 0 \circ x^3, 9 \circ x^4, 7 \circ x^5\},$$

$$A_7 = A_7^{(1)} + A_7^{(2)} = \{3 \circ x^1, 5 \circ x^2, 5 \circ x^3, 3 \circ x^4, 0 \circ x^5\},$$

$$A_8 = A_8^{(1)} + A_8^{(2)} = \{0 \circ x^1, 1 \circ x^2, 3 \circ x^3, 7 \circ x^4, 5 \circ x^5\}.$$

Table 3. Data table F[#]

$A \setminus K$	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8
A_1	4	5	4	5	4	5	4	5
	5	5	5	5	4	4	4	5
A_2	4	1	2	1	3	2	2	2
	3	2	1	1	4	3	3	2
A_3	1	1	3	1	4	1	1	4
	1	2	3	1	5	2	1	3
A_4	5	3	2	4	4	5	4	5
	4	4	3	5	4	5	3	4
A_5	4	4	4	4	4	5	4	4
	5	5	3	4	4	4	5	4
A_6	5	5	4	4	4	5	5	4
	4	5	4	4	4	4	5	5
A_7	4	1	2	3	3	3	1	2
	3	2	1	4	2	4	2	3
A_8	4	5	4	2	3	4	5	3
	5	4	5	3	4	5	4	4
A_9	3	2	3	1	3	3	2	2
	4	3	2	2	2	3	3	2
A_{10}	5	5	4	5	3	5	5	4
	3	4	3	4	2	4	2	4

Table 4. Data table G[#]

$A \setminus X$	x^1	x^2	x^3	x^4	x^5
A_1	0	0	0	4	4
	0	0	0	3	5
A_2	2	4	1	1	0
	2	2	3	1	0
A_3	5	0	1	2	0
	3	2	2	0	1
A_4	0	1	1	3	3
	0	0	2	4	2
A_5	0	0	0	7	1
	0	0	1	4	3
A_6	0	0	0	4	4
	0	0	0	5	3
A_7	2	2	3	1	0
	1	3	2	2	0
A_8	0	1	2	3	2
	0	0	1	4	3
A_9	1	3	4	0	0
	0	4	3	1	0
A_{10}	0	0	1	2	5
	0	2	2	4	0

Table 5 Objects–Attributes H shows objects A_1, \dots, A_{10} “generally” and the values of their attributes given by multisets of the form (22). Thus, for example, object A_1 is associated with the following multiset of estimates:

$$A_1 = \{0 \circ x_1^1, 0 \circ x_1^2, 0 \circ x_1^3, 1 \circ x_1^4, 1 \circ x_1^5; \\ 0 \circ x_2^1, 0 \circ x_2^2, 0 \circ x_2^3, 0 \circ x_2^4, 2 \circ x_2^5; \\ 0 \circ x_3^1, 0 \circ x_3^2, 0 \circ x_3^3, 1 \circ x_3^4, 1 \circ x_3^5; \\ 0 \circ x_4^1, 0 \circ x_4^2, 0 \circ x_4^3, 0 \circ x_4^4, 2 \circ x_4^5; \\ 0 \circ x_5^1, 0 \circ x_5^2, 0 \circ x_5^3, 2 \circ x_5^4, 0 \circ x_5^5; \\ 0 \circ x_6^1, 0 \circ x_6^2, 0 \circ x_6^3, 1 \circ x_6^4, 1 \circ x_6^5; \\ 0 \circ x_7^1, 0 \circ x_7^2, 0 \circ x_7^3, 2 \circ x_7^4, 0 \circ x_7^5; \\ 0 \circ x_8^1, 0 \circ x_8^2, 0 \circ x_8^3, 0 \circ x_8^4, 2 \circ x_8^5\}.$$

The table Objects–Attributes H[#], which represents individual versions of objects and their attributes, is rather cumbersome and is not given for brevity.

The proximity of objects in the space of attributes is characterized either by their difference or by the similarity of their properties. We will consider the multi-attribute object A_i represented by multiset A_i as a point of the metric space of multisets (\mathbf{A}, d) , $\mathbf{A} = \{A_1, \dots, A_m\}$, where d is one of the metrics (9)–(20). In practical problems, the difference between objects A_i and A_j is conveniently estimated by one of the metrics (17)–(20), which we write in the following form:

$$d_{Z_{1p}}(A_i, A_j) = [S_{ij}]^{1/p}; \\ d_{Z_{2p}}(A_i, A_j) = [S_{ij}/W]^{1/p}; \\ d_{Z_{3p}}(A_i, A_j) = [S_{ij}/M_{ij}]^{1/p}; \\ d_{Z_{4p}}(A_i, A_j) = [S_{ij}/N_{ij}]^{1/p}. \tag{23}$$

Here,

$$S_{ij} = m(A \Delta B) = \sum_{l=1}^n w_l |k_A(x^l) - k_B(x^l)| \\ = \sum_{l=1}^n w_l \sum_{e_l=1}^{h_l} |k_A(x_l^{e_l}) - k_B(x_l^{e_l})|; \\ L_{ij} = m(A \cap B) = \sum_{l=1}^n w_l \min |k_A(x^l), k_B(x^l)| \\ = \sum_{l=1}^n w_l \sum_{e_l=1}^{h_l} \min |k_A(x_l^{e_l}), k_B(x_l^{e_l})|; \\ M_{ij} = m(A \cup B) = \sum_{l=1}^n w_l \max |k_A(x^l), k_B(x^l)| \\ = \sum_{l=1}^n w_l \sum_{e_l=1}^{h_l} \max |k_A(x_l^{e_l}), k_B(x_l^{e_l})|;$$

Table 5. Data table H

$A \setminus X$	$x_1^1 x_1^2 x_1^3 x_1^4 x_1^5$	$x_2^1 x_2^2 x_2^3 x_2^4 x_2^5$	$x_3^1 x_3^2 x_3^3 x_3^4 x_3^5$	$x_4^1 x_4^2 x_4^3 x_4^4 x_4^5$	$x_5^1 x_5^2 x_5^3 x_5^4 x_5^5$	$x_6^1 x_6^2 x_6^3 x_6^4 x_6^5$	$x_7^1 x_7^2 x_7^3 x_7^4 x_7^5$	$x_8^1 x_8^2 x_8^3 x_8^4 x_8^5$
A_1	00011	00002	00011	00002	00020	00011	00020	00002
A_2	00110	11000	11000	20000	00110	01100	01100	02000
A_3	20000	11000	00200	20000	00011	11000	20000	00110
A_4	00011	00110	01100	00011	00020	00002	00110	00011
A_5	00011	00011	00110	00020	00020	00011	00011	00020
A_6	00011	00002	00020	00020	00020	00011	00002	00011
A_7	00110	11000	11000	00110	01100	00110	11000	01100
A_8	00011	00011	00011	01100	00110	00011	00011	00110
A_9	00110	01100	01100	11000	01100	00200	01100	00200
A_{10}	00101	00011	00110	00011	01100	00011	01001	00020

$$N_{ij} = m(A + B) = \sum_{l=1}^n w_l |k_A(x^l) + k_B(x^l)|$$

$$= \sum_{l=1}^n w_l \sum_{e_l=1}^{h_l} |k_A(x_l^{e_l}) + k_B(x_l^{e_l})|;$$

$$W = m(Z) = \sum_{l=1}^n w_l \max_{A \in \mathcal{A}} k_A(x^l)$$

$$= \sum_{l=1}^n w_l \sum_{e_l=1}^{h_l} \max_{A \in \mathcal{A}} k_A(x_l^{e_l}).$$

The measure of the multiset is given as a linear combination (6) of multiplicity functions: $m(A) = \sum_{l=1}^n w_l k_A(x^l) = \sum_{l=1}^n w_l \sum_{e_l=1}^{h_l} k_A(x_l^{e_l})$, where $w_l > 0$ is the significance (weight) of element x_l or, which is the same, of attribute K_l , $l = 1, \dots, n$. For quantities L_{ij} , M_{ij} , N_{ij} , and S_{ij} , relations $N_{ij} = M_{ij} + L_{ij}$, $S_{ij} = M_{ij} - L_{ij}$ are satisfied.

The similarity between objects can be expressed by one of the following indicators:

$$s_0(A_i, A_j) = L_{ij}; \quad s_1(A_i, A_j) = L_{ij}/W;$$

$$s_2(A_i, A_j) = 1 - d_{21}(A_i, A_j) = 1 - (S_{ij}/W);$$

$$s_3(A_i, A_j) = 1 - d_{31}(A_i, A_j) = L_{ij}/M_{ij};$$

$$s_4(A_i, A_j) = 1 - d_{41}(A_i, A_j) = 2L_{ij}/N_{ij}.$$

Expressions s_0, s_1, s_2, s_3 , and s_4 (24) generalize for the case of multisets the known nonmetric indices of the similarity of objects such as, respectively, the measure of absolute similarity, the Russell–Rao similarity measure, the simple matching coefficient, the Jacquard coefficient or the Rogers–Tanimoto measure, and the Sørensen index [4, 18].

In the example, the closest in the space of multisets (A, d_{11}) are objects A_4 and A_8 , which lie at a distance

$$d_{11}(A_4, A_8) = |0-0| + |0-0| + |0-0| + |0-0| + |0-0| = 0.$$

The group D_{01} , which includes objects A_4 and A_8 , is associated with multiset C_{01} formed by adding the multisets:

$$C_{01} = A_4 + A_8 = \{0 \circ x^1, 1 \circ x^2, 3 \circ x^3, 7 \circ x^4, 5 \circ x^5\}$$

$$+ \{0 \circ x^1, 1 \circ x^2, 3 \circ x^3, 7 \circ x^4, 5 \circ x^5\}$$

$$= \{0 \circ x_1, 2 \circ x_2, 6 \circ x_3, 14 \circ x_4, 10 \circ x_5\}.$$

Two pairs of objects A_1 and A_6 , A_2 and A_7 are at the same distances:

$$d_{11}(A_1, A_6) = |0-0| + |0-0| + |0-0| + |7-9| + |9-7| = 4,$$

$$d_{11}(A_2, A_7) = |4-3| + |6-5| + |4-5| + |2-3| + |0-0| = 4.$$

Objects A_1 and A_6 form group D_{02} represented by multiset

$$C_{02} = A_1 + A_6 = \{0 \circ x^1, 0 \circ x^2, 0 \circ x^3, 7 \circ x^4, 9 \circ x^5\}$$

$$+ \{0 \circ x^1, 0 \circ x^2, 0 \circ x^3, 9 \circ x^4, 7 \circ x^5\}$$

$$= \{0 \circ x^1, 0 \circ x^2, 0 \circ x^3, 16 \circ x^4, 16 \circ x^5\},$$

and objects A_2 and A_7 form group D_{03} represented by multiset

$$C_{03} = A_2 + A_7 = \{4 \circ x^1, 6 \circ x^2, 4 \circ x^3, 2 \circ x^4, 0 \circ x^5\}$$

$$+ \{3 \circ x^1, 5 \circ x^2, 5 \circ x^3, 3 \circ x^4, 0 \circ x^5\}$$

$$= \{7 \circ x^1, 11 \circ x^2, 9 \circ x^3, 5 \circ x^4, 0 \circ x^5\}.$$

The suggested indices of dissimilarity (23) and similarity (24) between multi-attribute objects were used in the methods of ordering ARAMIS (Aggregation and Ranking of Alternatives nearby Multi-attribute Ideal Situations), cluster analysis CLAVA-HI (CLuster Analysis of Verbal Alternatives—HIerarchical) and CLAVA-NI (CLuster Analysis of Verbal Alternatives—Non-Ierarchical), and classification MASKA (Multi-Attribute Consistent Classification of Alternatives) for

objects with mismatched attribute values [13, 15]. Examples of practical application of these methods in the problems of group multicriterion decision-making were given in [14–17].

CONCLUSIONS

This paper describes new classes of spaces of finite, bounded, and measurable sets and multisets and new types of metrics that can be used in solving various problems of classification, sorting, and ranking of multi-attribute objects. Representation of multi-attribute objects using multisets and application of new types of indicators of proximity of objects make it possible to expand the range of problems, as well as to more easily find solutions to known problems.

APPENDIX

Let us verify that the function $d_{Z_{4p}}(A, B) = [m(A\Delta B)/m(A + B)]^{1/p}$ defined on σ -algebra $\mathbf{S}_\sigma(\mathbf{Z})$ of multisets is quasi-metric (symmetric almost everywhere) on space $(\mathbf{S}_\sigma(\mathbf{Z}), m)$ of multisets with a strongly σ -additive and completely σ -finite measure.

The validity and nonnegativity of function $d_{Z_{4p}}(A, B)$ follow from the definition of measure m as a real and nonnegative multiset function defined on σ -algebra $\mathbf{S}_\sigma(\mathbf{Z})$ of multisets. The axiom of symmetry (1^0) follows from the commutativity of the addition operations $A + B = B + A$ and symmetric difference $A\Delta B = B\Delta A$ of multisets.

Let us show using examples that the triangle inequality (3^0) can be both satisfied and not for function $d_{Z_{4p}}(A, B)$. For $C \subseteq A \cap B$, by the monotonicity property of a measure of a multiset, we have $m(C) \leq$

$m(A \cap B) \leq m(A), m(C) \leq m(A \cap B) \leq m(B)$. Then, taking the inequality $m(A\Delta B) \leq m[(A\Delta C) + (B\Delta C)] = m(A\Delta C) + m(C\Delta B)$ proved in [6, 7] into account we obtain:

$$\begin{aligned} & m(A\Delta B)/m(A + B) \\ & \leq [m(A\Delta C) + m(B\Delta C)]/[m(A) + m(B)] \\ & \leq m(A\Delta C)/m(A + C) + m(B\Delta C)/m(B + C). \end{aligned}$$

Therefore,

$$\begin{aligned} [d_{Z_{4p}}(A, B)]^p & = m(A\Delta B)/m(A + B) \\ & \leq m(A\Delta C)/m(A + C) + m(B\Delta C)/m(B + C) \\ & \leq [d_{Z_{4p}}(A, C)]^p + [d_{Z_{4p}}(B, C)]^p \\ & \leq [d_{Z_{4p}}(A, C) + d_{Z_{4p}}(B, C)]^p \end{aligned}$$

and

$$d_{Z_{4p}}(A, B) \leq d_{Z_{4p}}(A, C) + d_{Z_{4p}}(B, C).$$

On the other hand, for arbitrary multisets A, B , and C , from formulas (7)–(8), we have

$$\begin{aligned} m(A \cup B) & = m(A) + m(B - A) = m(B) + m(A - B), \\ m(A\Delta B) & = m(A - B) + m(B - A), \end{aligned}$$

$$\begin{aligned} m(A\Delta(A \cup B)) & = m(A \cup B) - m(A) = m(B - A), \\ m(B\Delta(A \cup B)) & = m(A \cup B) - m(B) \\ & = m(A - B). \end{aligned}$$

Then for $C = A \cup B$, taking into account the inequality $m(A \cap B) \leq m(A), m(B) \leq m(A \cup B) \leq m(A + B)$, which follows from (8), we obtain

$$\begin{aligned} [d_{Z_{4p}}(A, B)]^p & = m(A\Delta B)/m(A + B) \\ & = m(B - A)/[m(A) + m(B)] + m(A - B)/[m(B) + m(A)] \\ & \geq m(B - A)/[m(A) + m(B) + m(A - B)] + m(A - B)/[m(B) + m(A) + m(B - A)] \\ & = m(A\Delta(A \cup B))/[m(A) + m(A \cup B)] + m(B\Delta(A \cup B))/[m(B) + m(A \cup B)] \\ & = m(A\Delta C)/m(A + C) + m(B\Delta C)/m(B + C) = [d_{Z_{4p}}(A, C)]^p + [d_{Z_{4p}}(B, C)]^p. \end{aligned}$$

It can be that $d_{Z_{4p}}(A, B) \geq d_{Z_{4p}}(A, C) + d_{Z_{4p}}(B, C)$.

By the identity property of the symmetric difference of multisets $A\Delta B = \emptyset$ for $A = B$. Then, $m(A\Delta B) = m(\emptyset) = 0$ and $d_{Z_{4p}}(A, B) = 0$. At the same time, the condition $m(A\Delta B) = 0$ in the general case does not imply the equality of multisets A and B . Thus, only the

coincidence condition (4^0) holds for function $d_{Z_{4p}}(A, B)$. Additionally, for multisets that differ by a multiset of measure zero, condition $m(A\Delta B) = 0$ implies the m -equality of multisets A and B , and hence the identity axiom (2^0) also holds. Thus, function $d_{Z_{4p}}(A, B)$ is quasimetric and almost everywhere sym-

metric on σ -algebra $\mathcal{S}_\sigma(\mathcal{Z})$ of measurable multisets with a strongly σ -additive and completely σ -finite measure.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (projects nos. 15-07-04760, 16-07-01125, 16-29-12864, and 17-07-00512) and the Russian Academy of Sciences Presidium (the program of fundamental research “Intelligent Information Technologies and Systems”).

REFERENCES

1. Lyusternik, L.A. and Sobolev, V.I., *Elementy funktsional'nogo analiza* (Elements of Functional Analysis), Moscow: Nauka, 1965.
2. Kolmogorov, A.N. and Fomin, S.V., *Elementy teorii funktsiy i funktsional'nogo analiza* (Elements of the Theory of Functions and Functional Analysis), Moscow: Nauka, 1968.
3. Deza, M.M. and Laurent, M., *Geometry of Cuts and Metrics*, Berlin: Springer-Verlag, 1997.
4. Sneath, P.H.A. and Sokal, R.R., Numerical taxonomy, in *The Principles and Practice of Numerical Classification*, San Francisco: Freeman, 1973.
5. Hartigan, J.A., *Clustering Algorithms*, New York: Wiley, 1975.
6. Petrovsky, A.B., *Prostranstva mnozhestv i mul'timnozhestv* (Spaces of Sets and Multisets), Moscow: Editorial URSS, 2003.
7. Petrovsky, A.B., *Prostranstva izmerimyykh mnozhestv i mul'timnozhestv* (Spaces of Measurable Sets and Multisets), Moscow: Poli Print Servis, 2016.
8. Petrovsky, A.B., *On distances between sets invariant under space transformations, I Vsesoyuznoye soveshchaniye po statisticheskomu i diskretnomu analizu nechisl'noy informatsii, ekspertnym otsenkam i diskretnoy optimizatsii. Tezisy dokladov* (I All-Union Meeting on Statistical and Discrete Analysis of Non-Numerical Information, Expert Assessments, and Discrete Optimization. Proceedings), Moscow–Alma-Ata, 1981, pp. 50–51.
9. Petrovsky, A.B., An axiomatic approach to metrization of multiset space, in *Multiple Criteria Decision Making*, Tzeng, G.H., Wang, H.F., Wen, U.P., and Yu, P.L., Eds., New York: Springer-Verlag, 1994, pp. 129–140.
10. Petrovsky, A.B., Metric spaces of multisets, *Dokl. Akad. Nauk*, 1995, vol. 344, no. 2, pp. 175–177.
11. Petrovsky, A.B., Combinatorics of multisets, *Dokl. Akad. Nauk*, 2000, vol. 370, no. 6, pp. 750–753.
12. Petrovsky, A.B., Multisets as a model for representation of multi-attribute objects in decision-making and pattern recognition, *Iskusstv. Intell.*, 2002, no. 2, pp. 236–243.
13. Petrovsky, A., Group verbal decision analysis, in *Encyclopedia of Decision Making and Decision Support Technologies*, Adam, F. and Humphreys, P., Eds., Hershey, New York: IGI Global, 2008, vol. 1, pp. 418–425.
14. Petrovsky, A., Multi-attribute classification of credit cardholders: Multiset approach, *Int. J. Manage. Decis. Making*, 2006, vol. 7, nos. 2–3, pp. 66–179.
15. Petrovsky, A.B., Methods for the group classification of multi-attribute objects (part 1, part 2), *Sci. Tech. Inf. Process.*, 2010, vol. 37, no. 5, pp. 346–356, 357–368.
16. Petrovsky, A., Group multiple criteria decision making: Multiset approach, in *Recent Developments and New Directions in Soft Computing. Studies in Fuzziness and Soft-Computing*, Springer International Publishing, 2014, pp. 19–33.
17. Petrovsky, A.B. and Lobanov, V.N., Multi-criteria choice in the attribute space of large dimension: Multi-method technology PAKS-M, *Sci. Tech. Inf. Process.*, 2014, no. 3, pp. 92–104.
18. Syomkin, B.I. and Dvoychenkov, V.I., On the equivalence of measures of similarity and difference, in *Issledovaniye sistem* (Research of Systems), vol. 1: *Analiz slozhnykh sistem* (Analysis of Complex Systems), Vladivostok: Dal'nevost. Nauchn. Tsentr Akad. Nauk SSSR, 1973, pp. 95–104.
19. Kusters, W.A. and Laros, J.F.J., Metrics for multisets, *Research and Development in Intelligent Systems XXIV, Proceedings of AI-2007*, Springer-Verlag London Limited, 2008, pp. 294–303.

Translated by K. Lazarev