Elementary Theory of Similarities and Its Use in Biology and Geography

B. I. Semkin

Pacific Institute of Geography, Vladivostok, ul. Radio 7, 690041 Russia

Abstract—A system of axioms of similarity measures, which is defined based on families of finite descriptive sets, is considered. The concepts of the equivalence and coequivalence of similarity measures and dissimilarity measures are presented. Diversity measures of one or more descriptive sets are introduced using similarity measures. Illustrative examples of similarity and dissimilarity measures are taken from biology and geography.

Keywords: Absolute similarity measure, dissimilarity measure, descriptive set, ordinary finite sets, finite multisets, normalized descriptive sets, weights of descriptive sets.

DOI: 10.1134/S1054661812010300

INTRODUCTION

Similarity measures and dual to them dissimilarity measures are central to the theoretical basis of cluster analysis and pattern recognition techniques. The final result of the partition of objects into classes for a given partition algorithm depends crucially on their choice [5, 47]. We used the axiomatic approach to arrange an infinite set of similarity measures and to limit the use of inappropriate measures. In this context, we introduce the notion of the equivalence of similarity measures [15, 17], based on which the classes of equivalent measures are indicated.

The elementary theory of similarity considers only the description of objects that are represented by finite descriptive sets. The basis of the theory of similarity is the following sections:

1. Descriptive sets, relationships, operations and measures on their basis;
2. Axioms of similarity measures of two descriptive sets (an absolute measure of similarity, relative inclusion measures, the relative similarity measure;
3. Equivalence of similarity measures and classes of equivalence of measures;
4. Measures of similarity and dissimilarity;
5. Harmonization of similarity measures with the algorithms for organizing the descriptive sets;
6. Similarity measures and their relationship with measures of diversity;
7. Multiple similarity measures;
8. Classes of equivalent multiplace similarity measures.

The calculations of various indices and coefficients of similarity, which were made on the basis of measurements, were initially carried out approximately based on intuition and often lead to erroneous results. For more than 100 years, beginning with the work of Swiss botanist P. Jaccard [31], a number of different coefficients of similarity and dissimilarity has been suggested in biology and geography, many of which were erroneous. Extensive reviews on the subject often only provide an opportunity to trace the history of the introduction and the use of this or that coefficient [2–3, 6, 7, 21–30, 32, 33, 35–46, 48, 49, 51].

A comparative analysis of similarity coefficients has only been carried out on some hypothetical examples [26]. To remedy this situation, it is necessary to verify the formal grounds that allow us to calculate the coefficients of similarity. The measurements and calculations of a similar nature were applied to different objects and their combinations and their variety has forced the researchers to distract from the observed relations of individual objects and move to generalized formal concepts and theory.

DESCRIPTIVE SETS

The mathematical concept of a "descriptive set" is to explain the concept of "assemblage," which is widely used in biology and geography. A descriptive set is a finite set, each element of which is assigned a positive number (weight) [9]. Since the descriptive set is determined by the values of weights, then we will use an equivalent concept, i.e., descriptive collection, which is an ordered set of weights.

Relations of equality and inclusion are set based on descriptive sets for each of the pairs $a = \langle a_1, \ldots, a_r\rangle$ and $b = \langle b_1, \ldots, b_r\rangle$ as follows:

$$a = b \Leftrightarrow a_i = b_i (i = 1, \ldots, r)$$
$$a \le b \Leftrightarrow a_i \le b_i (i = 1, \ldots, r)$$

Received August 8, 2011

and operations of conjunction, disjunction, and difference are written as follows:

$$a \wedge b = (\min(a_1, b_1), \ldots, \min(a_r, b_r));$$

$$a \vee b = (\max(a_1, b_1), \ldots, \max(a_r, b_r));$$

$$a \backslash b = (a_1 - \min(a_1, b_1), \ldots, a_r - \min(a_r, b_r)).$$

The weight of the descriptive vector (set) is the sum of weights of its components as follows:

$$m(a) = a_1 + \ldots + a_r,$$

$$m(\theta) = 0 \quad \theta = \underbrace{\langle 0, \ldots, 0 \rangle}_{r}.$$

Then, the following weights of the descriptive sets are often used:

$$m(a \wedge b) = \sum_{i=1}^{r} \min(a_i, b_i);$$

$$m(a \vee b) = \sum_{i=1}^{r} \max(a_i, b_i);$$

$$m(a \backslash b) = m(a) - m(a \wedge b)$$

$$= \sum_{i=1}^{r} a_i - \sum_{i=1}^{r} \min(a_i, b_i).$$

Five types of weights of the descriptive sets are of particular note:

1. Ordinary finite sets (SFVs) $a_i \in \{0, 1\}$ $(i = 1, \ldots, r)$.

2. Finite multisets (FMVs) $a_i \in \{0, 1, 2, \ldots, n\}(i = 1, \ldots, r)$ [8, 20].

3. Weighted sets (DWVs) $a_i \geq 0$, $(i = 1, \ldots, r)$ [9].

4. Normalized descriptive vectors by the components (NDVCs) $0 \leq a_i \leq 1$, $(i = 1, \ldots, r)$.

5. normalized overall descriptive vectors (NDVWs) $0 \leq a_i \leq 1$, $a_1 + \ldots + a_r$, $(i = 1, \ldots, r)$.

For each of these types of descriptive vectors, the considered similarity measures have some specific features. For this reason, in the future, we will specify types of sets of descriptive vectors that run these or other properties of similarity measures.

SYSTEMS OF AXIOMS OF ABSOLUTE SIMILARITY MEASURES

The absolute similarity measure of two descriptive sets a and $b(J(a, b))$ defined on set E of all considered types of weights of descriptive sets (1–5) is determined by the following system of axioms of the first group [18]:

ASM 1. $J(a, b) \geq 0$; $a, b \in E$;

ASM 2. $J(a, b) = 0 \Leftrightarrow a \wedge b = \theta$; $a, b \in E$;

ASM 3. $J(a, b) = J(b, a)$; $a, b \in E$;

ASM 4. $J(a, b) \leq J(a, a)$; $a, b \in E$;

ASM 5. $J(a, b) = J(a, a) \Leftrightarrow a \leq b$; $a, b \in E$;

ASM 6. $J(\lambda a, \lambda b) = \lambda J(a, b)$, $\lambda > 0$; $a, b \in E$;

ASM 7. $J(a, b) + J(b, c) \leq J(a, c) + J(b, b)$; $a, b, c \in E$.

The system of axioms of absolute similarity measures is consistent. For example, measure $J(a, b) = m(a \wedge b)$ satisfies the given axioms.

The second group of axioms is as follows:

ASM 8. The absolute similarity measures of two descriptive sets do not depend on other descriptive sets of multitude E.

ASM 9. The absolute similarity measures are only defined on descriptive sets, each component of which is measured in the same ratio scale.

RELATIVE MEASURE OF INCLUSION

The proposed system of axioms of absolute similarity measures allows us to construct different relative measures, in particular measures of the inclusion of one descriptive set into another. The measures of inclusion are determined by more than just the type of NDVW.

Corollary fact from axioms. It follows from ASM1 and ASM2 that $0 \leq \dfrac{J(a, b)}{J(a, a)} \leq 1$ and it follows from ASM3 and ASM4 that $J(a, b) \leq J(b, b)$. Because of the symmetry, we have $0 \leq \dfrac{J(a, b)}{J(b, b)} \leq 1$.

Let us denote the received ratio in the following form:

$$K(b; a) = \frac{J(a, b)}{J(a, a)},$$

$$K(a; b) = \frac{J(a, b)}{J(b, b)}.$$

Let us refer to these as the measures of inclusion of descriptive set a into b and b into a, respectively. Based on ASM1, ASM2, ASM5, and ASM6, we have the following properties $K(b; a)$:

RIM 1. $0 \leq K(b; a) \leq 1$ (limitedness);

RIM 2. $K(b; a) = 0 \Leftrightarrow a \wedge b = \theta$ (minimum inclusion);

RIM 3. $K(b; a) = 1 \Leftrightarrow a \leq b$ (maximum inclusion);

RIM 4. $K(\lambda b; \lambda a) = K(b; a)$ (homogeneity).

For example, it is obvious that measure $K(b; a) = \dfrac{m(a \wedge b)}{m(a)}$ satisfies the above properties. Measures of inclusion on NDVW-type vectors are meaningless because $m(a) = m(b) = 1$. The properties of RIM1–RIM 4 for measures of inclusion can be taken as a system of axioms of measures of inclusion for the descriptive normalized vectors. ASM8 axiom can be written as:

RIM 5. $K(b; a) = f[m(a), m(b), m(a \wedge b)]$ and $K(a; b) = g[m(a), m(b), m(a \wedge b)]$, where f and g are any continuous functions of three variables on the sets of numbers.

It follows from ASM4 that

$$J(a, b) \le J(a, a) \le (1+\tau)J(a, a) - \tau J(a, b)\tau \ge 0.$$

We obtain the continuum of measures of inclusion that are organized by parameter τ as follows:

$$K_\tau(b; a) = \frac{J(a, b)}{(1+\tau)J(a, a) - \tau J(a, b)},$$

$$K_0(b; a) = \frac{J(a, b)}{J(a, a)}.$$

Because of the symmetry we get:

$$K_\tau(a; b) = \frac{J(a, b)}{(1+\tau)J(b, b) - \tau J(a, b)},$$

$$K_0(a; b) = \frac{J(a, b)}{J(b, b)}.$$

In biology and geography, the following measures of inclusion are mainly used:

$$K_0(a; b) = \frac{\sum_{i=1}^{r}\min(a_i, b_i)}{\sum_{i=1}^{r} b_i},$$

$$K_0(b; a) = \frac{\sum_{i=1}^{r}\min(a_i, b_i)}{\sum_{i=1}^{r} a_i},$$

as well as dual measures of exclusion:

$$F_0(a; b) = 1 - K_0(a; b),$$

$$F_0(b; a) = 1 - K_0(b; a).$$

RELATIVE SIMILARITY MEASURES

From the proposed system of axioms for absolute similarity measures a number of consequences can also be derived which are necessary for the construction of relative similarity measures. For example, based on two inequalities $J(a, b) \le J(a, a)$ and $J(a, b) \le J(b, b)$ we obtain the following relations:

$$0 \le \frac{2J(a, b)}{J(a, a) + J(b, b)} \le 1;$$

$$0 \le \frac{J(a, b)}{J(a, a) + J(b, b) - J(a, b)} \le 1.$$

We will refer to these relations as "relative similarity measures" between the descriptive sets a and b and denote them as $K_0(a, b)$ and $K_1(a, b)$, respectively.

From the axioms of absolute similarity measures for these measures, the following properties, which can be taken as axioms, are fair:

RSM 1. $0 \le K(a, b) \le 1$ (axiom of limitedness);

RSM 2. $K(a, b) = K(b, a)$ (axiom of symmetry);

RSM 3. $K(a, b) = 0 \Leftrightarrow a \wedge b = \theta$ (axiom of minimum similarity);

RSM 4. $K(a, b) = 1 \Leftrightarrow a = b$ (axiom of maximum similarity);

RSM 5. $K(\lambda a, \lambda b) = K(a, b)$ (axiom of homogeneity).

This system of axioms is not contradictory to it, e.g., the relative similarity measure $K(a, b) = \frac{2m(a \wedge b)}{m(a) + m(b)}$ satisfies it.

Based on the system of axioms ASM, a number of consequences can be obtained.

Statement 1.

Function

$$K_\tau(a, b) = \frac{J(a, b)}{(1+\tau)[J(a, a) + J(b, b)] - \tau J(a, b)},$$
$$-1 \le \tau \le +\infty,$$

satisfies the properties of the relative similarity measures (RSMs).

The continuum of relative similarity measures that are organized by parameter τ tau, can be briefly written as

$$K_\tau(a, b) = \frac{K_0(a, b)}{(1+\tau) - \tau K_0(a, b)},$$

where $K_0(a, b) = \frac{2m(a \wedge b)}{m(a) + m(b)}$, $-1 \le \tau \le +\infty$.

Statement 2.

The two-parameter function

$$K_{\tau;\eta}(a, b) = \left[\frac{K_\tau^\eta(a; b) + K_\tau^\eta(a; b)}{2}\right]^{1/\eta},$$

where $K_\tau(a, b) = \frac{K_0(a; b)}{(1+\tau) - \tau K_0(a; b)}$; $K_\tau(b, a) = \frac{K_0(b; a)}{(1+\tau) - \tau K_0(b; a)}$; $K_0(a, b) = \frac{m(a \wedge b)}{m(b)}$; $K_0(b, a) = \frac{m(a \wedge b)}{m(a)}$; $-1 \le \tau \le +\infty$; $-\infty \le \eta \le +\infty$ satisfies the system of axioms RSM.

Let us mark some specific measures as follows:

$$K_{\tau;0}(a, b) = \sqrt{K_\tau(a; b)K_\tau(b; a)};$$

$$K_{\tau;-\infty}(a, b) = \min[K_\tau(a; b), K_\tau(b; a)];$$

$$K_{\tau;+\infty}(a, b) = \max[K_\tau(a; b), K_\tau(b; a)].$$

Note that the measures of inclusion disclose the relationship between the compared objects more fully than the widely used similarity measures, and similarity measures are derived from measures of the inclusion and give them the average estimation.

METRIC SIMILARITY MEASURES

Similarity measures, which satisfy the inequality of four RSM measures six, are called metric similarity measures.

RSM 6. $K(a, b) + K(b, c) \le K(a, c) + K(b, b)$, $K(b, b) = 1$ $a, b, c \in$ E.

For example, $K(a, b) = \dfrac{m(a \wedge b)}{m(a) + m(b) - m(a \wedge b)}$ is a metric similarity measure.

If axiom RSM 4 is not fully performed for a measure (i.e., from condition $K(a, b) = 1 \Rightarrow a \ne b, \forall a, b \in E$, this measure is called a quasi-measure of similarity. For example, the $K_{0;\,-\infty}(a, b) = \dfrac{m(a \wedge b)}{\min(m(a), m(b))}$.

FRACTIONAL-LINEAR RELATIVE SIMILARITY MEASURES

Axiom ASM 8 can be written for the case of relative similarity measures in the following form:

RSM 7. Similarity measure $K(a, b)$ is only determined by $m(a)$, $m(b)$, and $m(a \wedge b)$, i.e.,

$$K(a, b) = f[m(a), m(b), m(a \wedge b)],$$

where f is a continuous function of three arguments.

Statement 3.

Let

$$K(a, b) = f[m(a), m(b), m(a \wedge b)]$$
$$= \frac{\alpha_1 m(a) + \alpha_2 m(b) + \alpha_3 m(a \wedge b) + \alpha_4}{\beta_1 m(a) + \beta_2 m(b) + \beta_3 m(a \wedge b) + \beta_4},$$

where α_i and β_i (for $i = 1, ..., 4$) are unknown constants. Then the relative similarity measure $K(a, b)$ satisfies the axioms of RSM with the following constants: $\alpha_1 = \alpha_2 = \alpha_4 = \beta_4 = 0$, $\alpha_3 = 2$, $\beta_1 = \beta_2 = 1 - \tau$, $\beta_3 = -2\tau$, $-1 \le \tau \le +\infty$ or

$$K_\tau(a, b) = \frac{K_0(a, b)}{(1+\tau) - \tau K_0(a, b)},$$

$$K_0(a, b) = \frac{2m(a \wedge b)}{m(a) + m(b)}, \quad -1 \le \tau \le +\infty.$$

RELATIVE DISSIMILARITY MEASURES

The relative dissimilarity measure is the addition of relative similarity measure up to the unit as follows:

$$F(a, b) = 1 - K(a, b).$$

The set of properties of dissimilarity measures (RDM) follows from the axiom system RSM as shown below:

RDM 1. $0 \le F(a, b) \le 1$ (limitedness);

RDM 2. $F(a, b) = F(b, a)$ (symmetry);

RDM 3. $F(a, b) = 0 \Leftrightarrow a = b$ (minimum difference);

RDM 4. $F(a, b) = 1 \Leftrightarrow a \wedge b = \theta$ (maximum difference);

RDM 5. $F(\lambda a, \lambda b) = F(a, b)$ (homogeneity).

If the property RDM3 is only partially performed, i.e., when $F(a, b) = 0 \Leftrightarrow a \ne b$, this measure is called the measure of quasi-dissimilarity (denote this property as RKDM3).

In the case of implementing dissimilarity measures, the terms of the triangle inequality as follows:

RDM 6. $F(a, b) + F(b, c) \ge F(a, c)$, $a, b, c \in E$;

the dissimilarity measure is called distance.

When the properties RKDM3 and RDM6 are performed, the measure is called quasi-distance.

EQUIVALENCE OF MEASURES

Before applying the axiomatic approach to the introduction of measures of similarity and dissimilarity, reviews of measures of proximity had a primarily historical character. The author, who first used a certain rate, as well as those who introduced modifications to it for their own purposes, were mentioned; usually anywhere from a dozen to several dozen different coefficients of proximity were cited.

When using an axiomatic approach to the introduction of measures of similarity and dissimilarity, a continual set of different similarity measures ordered by certain parameters appeared. There was a problem of treatment with a continuum of measures, which was qualitatively different from a separately identified index.

The first works on the introduction of the equivalence of similarity measures [4, 13, 17, 45] allowed one to give a strict definition of the concept of equivalence of similarity measures [15].

Definition 1. Two similarity measures $K(a, b)$ and $K'(a, b)$ are equivalent to the family of descriptive set E if for every pair $a, b \in E$ there is a monotonically increasing function φ: $K = \varphi(K')$, and $\varphi(0) = 0$, $\varphi(1) = 1$.

Example 1. Consider the following class of equivalent measures on the sets of SFV, FMV, DWV, and NDVC:

$$K_{\tau;\,-1}(a, b) = \frac{K_{0;\,-1}(a, b)}{1 + \tau - \tau K_{0;\,-1}(a, b)},$$

$$K_{0;\,-1}(a, b) = \frac{2m(a \wedge b)}{m(a) + m(b)}, \quad -1 < \tau < +\infty.$$

Example 2. Inequivalent measures on vectors SFV, FMV, DWV, and NDVC: $K_{0;\,-1}(a, b)$ and $K_{0;\,1}(a, b) = \dfrac{1}{2} m(a \wedge b)\left[\dfrac{1}{m(a)} + \dfrac{1}{m(b)}\right]$.

Example 3. A class of equivalent measures on descriptive vectors NDVW: $K(a, b) = \varphi(m(a \wedge b))$, where φ is a monotonically increasing function, and $\varphi(0) = 0$, $\varphi(1) = 1$.

Typically, the simplest representative is usually chosen from this class of measures as follows [4]:

$$J(a, b) = m(a \wedge b) = \sum_{i=1}^{r} \min(a_i, b_i),$$

$$\sum_{i=1}^{r} a_i = \sum_{i=1}^{r} b_i = 1.$$

The following definition is equivalent to the definition of equivalence [17]:

Definition 2. If, for any descriptive vectors, a, b, c, $d \in E$, condition

$$K(a, b) \geq K(c, d) \Leftrightarrow K'(a, b) \geq K'(c, d)$$

is done. Then the measures K and K' are equivalent by E.

Note that the equation is either not achieved or performed in both inequalities simultaneously.

A corollary fact. In Definitions 1 and 2, if we replace the similarity measure K and K' to their dual dissimilarity measures F and F', then we obtain a definition of equivalence for the dissimilarity measures.

Definition 3. If $K(a, b)$ and $F(a, b)$ are such that there is a strictly monotonically decreasing function ψ ($\psi(0) = 0$, $\psi(1) = 1$), and $K(a, b) = \psi F(a, b)$, then $K(a, b)$ and $F(a, b)$ are called coequivalent [17].

We denote the equivalence using the sign $\|$ and the coequivalence using the sign $\perp$. Then. the following statement is fair:

Statement 4.

If $K_1 \| K_2$ and $K_2 \perp F$, then $K_1 \perp F$.

If $K_1 \perp F$ and $F \perp K_2$, then $K_1 \| K_2$.

The proof is given in [17]. For example, if $K_0(a, b) \| K_1(a, b)$ and $K_1(a, b) \perp F_1(a, b)$, then $K_0(a, b) \perp F_1(a, b)$.

Statement 5. Equivalent relative dissimilarity measures $F_{\tau; -1}(a, b)$ defined on descriptive vectors, such as SFV, FMV, DWV and NDVC, are the distances at $\tau \geq 1$. If $F_\tau(a, b)$ are defined on descriptive vectors (NDVW), they are distances at $\tau \geq 0$.

The first part of the proof of this Statement can be found in [17]. Let us prove the second part of the statement.

Proof. Suppose $m(a) = m(b) = 1$, then $F_{\tau; -1}(a, b) = \dfrac{1 - m(a \wedge b)}{1 - \varphi_\tau + \varphi_\tau m(a \wedge b)}$, $-1 < \tau < +\infty$. It is obvious that $1 - m(a \wedge b) = \rho(a, b)$ is the distance because, if $m(a) = m(b) = 1$, the Hamming distance is $2(1 - m(a \wedge b) = 2\rho(a, b)$. Consequently, $F_{\tau; -1}(a, b) = \dfrac{\rho(a, b)}{1 - \varphi_\tau + \varphi_\tau \rho(a, b)}$ is the distance if $1 - \varphi_\tau > 0$ and $\varphi_\tau \geq 0$ or $\tau \geq 0$.

Statement 6. Equivalent relative dissimilarity measures.

$$F_{\tau; -\infty}(a, b) = \frac{m(a) + m(b) - 2m(a \wedge b) + |m(a) - m(b)|}{m(a) + m(b) - 2\varphi_\tau m(a \wedge b) + |m(a) - m(b)|},$$

$\varphi_\tau = \dfrac{\tau}{1 + \tau}$, $-1 < \tau < +\infty$ for any descriptive vectors a, b, $c \in E$ are distances if $\tau \geq 0$.

The Proof. Previously [10] it has been proved that the relative dissimilarity measure $F_{0; -\infty}(a, b) = \dfrac{m(a) + m(b) - 2m(a \wedge b) + |m(a) - m(b)|}{m(a) + m(b) + |m(a) - m(b)|}$ is Yurtsev distance. Let us express $F_{\tau; -\infty}(a, b)$ through $F_{0; -\infty}(a, b)$:

$$F_{\tau; -\infty}(a, b) = \frac{F_{0; -\infty}(a, b)}{1 - \varphi_\tau + \varphi_\tau F_{0; -\infty}(a, b)}.$$

It is obvious that the relative dissimilarity measures at $0 \leq \varphi_\tau \leq 1$ ($\tau \geq 0$) are distances.

For example,

$$F_{1; -\infty}(a, b) = \frac{m(a) + m(b) - 2m(a \wedge b) + |m(a) - m(b)|}{m(a) + m(b) - m(a \wedge b) + |m(a) - m(b)|} = \frac{m(a \vee b) - m(a \wedge b) + |m(a) - m(b)|}{m(a \vee b) + |m(a) - m(b)|}$$

is the distance.

DEFINITION OF DIVERSITY MEASURES USING SIMILARITY MEASURES

Consider a set of descriptive vectors of type NDVW. Let p, $q \in E$ and have the following form:

$$p = \langle p_1, \ldots, p_r \rangle, \quad \sum_{i=1}^{r} p_i = 1 \ (i = 1, \ldots, r), \ p_i \geq 0;$$

$$q = \langle q_1, \ldots, q_r \rangle, \quad \sum_{i=1}^{r} q_i = 1 \ (i = 1, \ldots, r), \ q_i \geq 0.$$

From the set of equivalent measures, which are defined based on E, let us take the simplest measure

$$J(p, q) = \sum_{i=1}^{r} \min(p_i, q_i).$$

We define the diversity measure of descriptive vector $p = \langle p_1, \ldots, p_r \rangle$ as [11, 19]:

$$R(p) = \sum_{i=1}^{r} \min\left(p_i, \frac{1}{r}\right).$$

The dual measure of concentration is determined by the relative dissimilarity measure [1]

$$Q(p) = 1 - R(p) = \frac{1}{2} \sum_{i=1}^{r} \left| p_i, \frac{1}{r} \right|.$$

MULTIPLACE SIMILARITY MEASURES

Multiplace similarity measures are widely used in biology and geography to measure β-diversity [9, 15, 25, 34, 50]. An axiomatic approach to multiplace similarity measures and their dual dissimilarity measures is proposed in [9, 12, 14].

The following is a system of axioms for assessing the degree of similarity of series of n descriptive vectors $\{a^i\}$ $(i = 1, \ldots, n)$:

MSM 1. $0 \le K(a^{(1)}, \ldots, a^{(n)}) \le 1$ (axiom of limitations);

MSM 2. $K(a^{(1)}, \ldots, a^{(n)}) = K(a^{(i_1)}, \ldots, a^{(i_n)})$ (axiom of symmetry), where $i_1, \ldots, i_n$ is any permutation of the indices $1, \ldots, n$;

MSM 3.

$$K(a^{(1)}, \ldots, a^{(n)}) = 0$$
$$\Leftrightarrow \underbrace{a^{(1)} \wedge a^{(2)} = \Theta, \ldots, a^{(n-1)} \wedge a^{(n)} = \Theta}_{\frac{n(n-1)}{2}}$$

(axiom of minimum similarity);

MSM 4. $K(a^{(1)}, \ldots, a^{(n)}) = 1 \Leftrightarrow a^{(1)} = \ldots = a^{(n)}$ (axiom of maximum similarity);

MSM 5. $K(\lambda a^{(1)}, \ldots, \lambda a^{(n)}) = K(a^{(1)}, \ldots, a^{(n)})$, $\lambda > 0$ (axiom of homogeneity).

CONCLUSIONS

The use of the axiomatic approach to constructing an elementary theory of similarity has led, on one hand, to the construction of a continuum of measures of similarity and dissimilarity and, on the other hand, to the separation of classes of equivalent measures, which gives the same organization of descriptive vectors and allows one to use only one the simplest measure of each class.

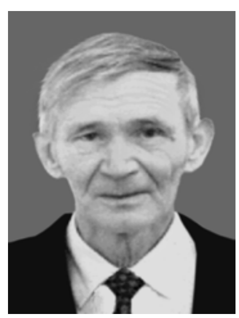
The development for similarity measures defined based on NDVWs and the class of linear fractional similarity measures ($K_{\tau; -\infty}$) can now be considered fully completed. These two classes of similarity measures and their dual dissimilarity measures are most often used in biology and geography. In the future, the elementary theory of similarity can be extended by a homomorphism on the other objects and measures, such us dependence and compatibility [16].

REFERENCES

1. R. L. Ackoff and F. E. Emery, *On Purposeful Systems* (Tavistock, London, 1972; Sov. Radio, Moscow, 1974).
2. V. D. Aleksandrova, *Flora Classification. Review of Classification Principles and Classification Systems in Different Geobotanical Schools* (Nauka, Leningrad, 1969) [in Russian].
3. V. I. Vasilevich, *Statistical Methods in Geobotanics* (Nauka, Leningrad, 1969) [in Russian]; V. B. Kovalevskaya, I. B. Pogozhev, and A. P. Pogozheva (Kusurgasheva), "Numerical Methods for Estimating the Proximity Degree of Archeology Monuments by Percentage of Mass Material," Sov. Arkheol., No. 3, 26–39 (1970).
4. V. N. Kotov, *Measuring Theory in Biological Researches* (Naukova dumka, Kiev, 1985) [in Russian].
5. B. M. Mirkin and G. S. Rozenberg, *Thesaurus of Modern Phytocenology* (Nauka, Moscow, 1983) [in Russian].
6. Ya. A. Pesenko, *Principles and Methods of Numerical Analysis in Fauna Research* (Nauka, Moscow, 1982) [in Russian].
7. A. B. Petrovskii, "Multisets Metrical Spaces," Dokl. Akad. Nauk **344** (2), 175–177 (1995).
8. B. I. Semkin, "Descriptive Sets and Their Application," in *Systems Research*, Vol. 1: *Complex Systems Analysis* (DVNTs AN SSSR, Vladivostok, 1973), pp. 83–94.
9. B. I. Semkin, "Numerical Indexes for Estimating the Floral Unilateral Constraints Offered by B. A. Yurtsev," Botan. Zh. **92** (4) 114–127 (2007).
10. B. I. Semkin, "On Theoretical Set Methods for Researching the Floral Communities," in *Proc. 5th Meeting of All-Union Botanical Society* (Kiev, 1973), pp. 210–211.
11. B. I. Semkin, "On Axiomatic Approach to Detect Disparity Measures and Quasidifference for Sets Families," in *Information Methods in Control Systems for Measurements and Inspection* (DVNTs AN SSSR, Vladivostok, 1972), Vol. 1, pp. 23–26.
12. B. I. Semkin, "Disparity Measures Equivalence and Hierarchical Classification of Multi-Dimensional Data," in *Hierarchical Classification Generations in Geographical Ecology and Systematics* (DVNTs AN SSSR, Vladivostok, 1979), pp. 97–112.
13. B. I. Semkin and M. V. Gorshkov, "Axiomatic Introduction of Measures for Affinity, Difference, Compatibility and Regularity for Components of Biological Variety for Multi-Dimensional Case," Vestn. KrasGAU, No. 12, 18–24 (2009).
14. B. I. Semkin and M. V. Gorshkov, "The Way to Estimate Affinity and Difference in Floral and Phytocenoscal Descriptions," in *Proc. Komarov Readings* (2010), Issue 57, pp. 203–220.
15. B. I. Semkin and M. V. Gorshkov, "A Set of Axioms for Symmetrical Functions for Two Variables and Measures for Affinity, Difference, Compatibility, and Regularity for Components of Biological Variety," Vestn. TGEU, No. 4. 31–46 (2008).
16. B. I. Semkin and V. I. Dvoichenkov, "Affinity and Difference Equivalence," in *Systems Research*, Vol. 1: *Complex Systems Analysis* (DVNTs AN SSSR, Vladivostok, 1973), pp. 95–104.
17. B. I. Semkin, A. P. Oreshko, and M. V. Gorshkov, "How to Use the Bioinformation Technologies in Comparative Floristics. I. Scheme-Goal Approach. Absolute Measures for Affinity and Difference," Byull. BSI DVO RAN (Vladivostok, 2009), Issue 3, pp. 102–111, Available from: http://www.botsad.ru/journal/number3/02.
18. B. I. Semkin, N. G. Klochkova, I. S. Gusarova, and M. V. Gorshkov, "Discreteness and Continuity for

Flora of Weed-Macrophyte of Far East Russian Seas. III. Taxonomic Spectra," Izv. TINRO **163**, 217–227 (2010).

19. M. Aigner, *Combinatorial Theory* (Springer, Berlin, 1997).
20. J. P. Cancella da Fonseca, "L'outie statistique en biologie du sol," V. Indices de diversite specifique. Rev. d'ecol et biol. du sol **6** (1), 1–30 (1969).
21. A. Chao, W. H. Hwang, Y. C. Chen, and C. Y. Kuo, "Estimating the Number of Shared Species in Two Communities," Statistica Sinica, No. 10, 227–246 (2000).
22. A. H. Cheetam and J. E. Hazel, "Binary (Presence-Absence) Similarity Coefficients," J. Paleontol. **43** (5), 1130–1136 (1969).
23. P. Y. N. Digby and R. A. Kempton, *Multivariate Analysis of Ecological Communities* (Chapman and Hall, London, New York, 1987).
24. O. H. Diserud and F. Odegaard, "A Multiple-Site Similarity Measure," Biol. Lett., No. 3, 20–22 (2007).
25. D. W. Goodall, "Sample Similarity and Species Correlation," in *Handbook of Vegetation Science*, Part 5: *Ordination and Classification of Vegetation* (The Hague, 1973), pp. 107–156.
26. P. Greig-Smith, *Quantitative Plant Ecology* (Butterworths, London, 1957).
27. P. Greig-Smith, *Quantitative Plant Ecology*, 2nd ed. (Butterworths, London, 1964).
28. P. Greig-Smith, *Quantitative Plant Ecology*, 3rd ed. (Butterworths, London, 1983).
29. M. O. Hill, "Diversity and Evenness: A Unifying Notation and Its Consequence," Ecology **54** (2), 427–432 (1973).
30. P. Jaccard, "Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines," Bull. Soc. Vaudoise sci. Natur. **37** (140), 241–272 (1901).
31. K. A. Kershaw, *Quantitative and Dynamics Plant Ecology* (Arnold, London, 1964).
32. K. A. Kershaw and J. H. H. Looney, *Quantitative and Dynamics Plant Ecology*, 3rd ed. (Arnold, London, 1985).
33. L. F. Koch, "Index of Biotal Dispersity," Ecology **38** (1), 145–148 (1957).
34. P. Koleff, K. J. Gaston, and J. J. Lennon, "Measuring Beta Diversity for Presence-Absence Data," J. Animal Ecol. **72**, 367–382 (2003).
35. J. M. Lambert and M. B. Dale, "The Use of Statistics in Phytosociology," Adv. Ecol. Res. **2**, 59–99 (London, New York, 1964).
36. L. Legendre and P. Legendre, *Numerical Ecology* (Elsevier, Amsterdam, 1983).
37. A. E. Magurran, *Ecological Diversity and Its Measurement* (Princeton Univ. Press, New Jersey, 1988).
38. A. E. Magurran, *Measuring Biological Diversity* (Blackwell Publ., Oxford, UK, 2004).
39. H. J. Oosting, *The Study of Plant Communities* (W. H. Freeman and Co., San Francisco, 1953).
40. J. A. Peters, "A Computer Program for Calculating Degree of Biogeographical Resemblance between Areas," Syst. Zool. **17** (1), 64–69 (1968).
41. E. C. Pielou, *Ecological Diversity* (Wiley-Intersci., New York, 1975).
42. E. C. Pielou, *An Introduction to Mathematical Ecology* (Wiley-Intersci., New York, 1969).
43. E. C. Pielou, *Mathematical Ecology*, 2nd ed. (Willey-Intersci., J. Wiley and Sons, New York–Chichester–Brisbane–Toronto, 1977).
44. G. Roux and M. Roux, "A propos du quelques methodes de classiffication en phytosociologie," Rev. Stat. Appl. **15** (2), 59–72 (1967).
45. J. M. Savage, "Evolution of a Peninsular Herpetofauna," Syst. Zool. **9**, 184–212 (1960).
46. R. Sibson, "A Model for Taxonomy. II," Math. Biosci., No. 6, 405–430 (1970).
47. P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy: The Principles and Practices of Numerical Classification* (Freeman, San-Francisco, 1973).
48. R. R. Sokal and P. H. A. Sneath, *Principles of Numerical Taxonomy* (Freeman, San Francisco, London, 1963).
49. R. H. Whittaker, *Communities and Ecosystems* (Macmillan, New York, London, 1970).
50. W. T. Williams and M. B. Dale, "Fundamental Problems in Numerical Taxonomy," Adv. Bot. Res. **2**, 35–68 (1965).



Boris Ivanovich Semkin. Born in 1938. Graduated from the Far Eastern Federal University, Department of Physics and Mathematics in 1962. Doctor of Biological Sciences, works at the Pacific Institute of Geography, Far East Branch of the Russian Academy of Sciences, Vladivostok. He is a lead researcher and professor at the Department of Ecology. His scientific interests similarity theory and algorithms for classifying descriptive sets.