Raport Badawczy
Research Report

RB/63/2013

On interaction between groups of objects – multiset approach

M. Krawczak, G. Szkatuła

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.:(+48) (22) 3810100

fax:(+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. dr hab. inż. Janusz Kacprzyk

Warszawa 2013

On interaction between groups of objects – multiset approach

Maciej Krawczak and Grażyna Szkatuła

Systems Research Institute, Polish Academy of Sciences, Newelska 6, Warsaw, Poland
{krawczak, szkatulg}@ibspan.waw.pl

Abstract: Properties of objects can be specified by a set of attributes which values can be symbolic. Defining a good measure of proximity (or remoteness) between objects is crucial importance in applied research like data mining or machine learning. Much attention has been devoted to continuous attributes while little attention to nominal attributes which seems to be much more difficult to handle. The proposed approach is based on the theory of multisets. In the paper we defined a group of objects with nominal description as a *K*-tuple of multisets (i.e., an ordered collection of multisets). There are introduced the following definitions: perturbation of multisets' for each attribute and a measure of groups' perturbations. The measure of perturbation is assumed to return a value from [0, 1], where 1 is interpreted as the most level of perturbation, while 0 is the lowest level of perturbation. In general these two measures are different, asymmetrical, so they should not be considered as the distance between the groups.

Keywords: Symbolic data analysis, Multisets, Measure of proximity, Groups of objects, Nominal attributes

1. Introduction

Defining a good distance measure between objects is of crucial importance in, for example, many classification and grouping algorithms. From the mathematical point of view, distance is defined as a quantitative degree showing how far apart two objects are. Synonyms for distance measure is dissimilarity between two groups. They are used to express the degree in which two objects are found to be different, usually on [0, 1] scale.

Every non empty subset of a finite set of objects is called a group. Note that the measure of distance between objects belonging to the groups can be defined as any measure appropriate to the type of grouped objects, such as quantitative data, ordinal, nominal, or encoded by the values 0 and 1. The adopted measure of distance has a great influence on obtained groups of different shapes. Most frequently used methods of calculating the distances between the groups are following:

- minimum distance between any two objects of each group;
- maximum distance between any two objects of each group;
- average distance between all pairs of objects in the two different groups,
- distance of average.

While a lot of work has been performed on continuous attributes, nominal attributes are more difficult to handle. Nominal data contains data with nominal attributes whose values neither have a natural ordering nor an inherent order. The variables of nominal data are measured by nominal scales. An attribute is nominal if it can take one of a finite number of possible values and, unlike ordinal attributes; these values bear no internal structure. An example is the attribute "taste", which may take the value of "salty", "sweet", "sour", "bitter" or "tasteless". When a nominal attribute can only take one of two possible values, it is usually called binary or dichotomous.

When the attributes are nominal, definitions of the similarity (or dissimilarity) measures become less trivial. Finding similarities between nominal objects by using common distance measures, which are used for processing numerical data, is not applicable here.

The proposed approach is based on the theory of the multisets. Classical set theory states that a set is a collection of distinct values. If repeating of any values is allowed in a set then such a set is called the *multiset* (sometimes also shortened to *mset* or *bag*). This way, the multiset is understood as a set with additional information about the multiplicity of occurring elements. Let us assume now that every subset of finite set V of nominal values, in which repetition of elements is significant, is called a *multiset*. The first time the term multisets was used by Dedekind in 1898. A complete survey of multisets theory can be found in many papers where several operations and their properties are investigated. A multiset can be expressed using different notations. An exemplary multiset containing one occurrence of a, three occurrences of b and two occurrence of c can be described by $\{a^1, b^3, c^2\}$ or $\{a,b,b,b,c,c\}$, or $\{a,b,c\}_{1,3,2}$, or $\{a1, b3, c2\}$, or $\{1/a, 3/b, 2/c\}$, or $\{(1,a),(3,b),(2,c)\}$ etc. or use square brackets, depending on adopted notation. In this paper we apply the sixth formalism, i.e., $\{(1,a),(3,b),(2,c)\}$.

In this paper, we defined the measure of perturbation of one multiset by another multiset. We defined a description of a group of objects as a K-tuple of multisets (an ordered collection of multisets). This concept was extended on all multisets within describing the considered groups; as a result we defined a measure of perturbation one group by another group. The idea of the measure of group's perturbation is based on a relation between two attributes' values represented as multisets, where each multiset belongs to different group's pair. Instead of considering dissimilarities between groups, we introduced a *measure of perturbation of one group by another group*. This measure defines changes of one group after adding a second group, and vice versa. It is interesting that this measure is not symmetric, it means a value of the measure of perturbation of a first group by a second group can be different then a value of the measure of perturbation of a second group by a first group. The measure of perturbation is assumed to return a value from [0, 1], where 1 is interpreted as the most level of perturbation, while 0 is the lowest level of perturbation. The measure is asymmetrical, so it should not be considered as the distance between the groups, but the diversity, assuming that one group is a group of the base.

The paper is organized as follows: in Section 2, we present preliminaries and basic definitions for multisets. In Section 3 we present description of the methodology, the measures of perturbation of one multiset by another multiset. In Section 4 we present the measures of interaction between groups of objects of nominal values.

2. Preliminaries and the basic definitions of multisets

From a practical point of view multisets are very useful structures arising in many areas of mathematics and computer science. In this section the basic definitions and notions of functions in multisets context are presented. These concepts will be extended to all multisets describing the considered groups of objects.

Let us consider a non-empty and finite set V of nominal values. For consecutive values labeled by letters of the alphabet, we get exemplary set as e.g. $V = \{a,b,c,d,e,f,g\}$; or when are labeled by the words, we get exemplary set as e.g. $V = \{\text{"salty"}, \text{"sweet"}, \text{"sour"}, \text{"bitter"}, \text{"tasteless"}\}$ or V= {"small", "medium", "large"}.

Definition 1. *Let us assume that the multiset S drawn from the ordinary set V can be represented by the ordered set of pairs:*

$$S = \{(k_S(v), v)\},\ \forall v \in V \tag{1}$$

where $k_S : V \to \mathrm{N}=\{0,1,2, \ldots\}$, $k_S(.)$ is called a *counting function* and the value $k_S(v)$ specifies the number of occurrences of the element $v \in V$ in the multiset S. Those elements which are not included in the multiset S have zero count.

A multiset S (1) drawn from the finite set V, $V = \{v_1, v_2, ..., v_L\}$, $v_{i+1} \neq v_i$, $\forall i \in \{1,2,..., L-1\}$, can be described by the set

$$S = \{(k_S(v_1), v_1), (k_S(v_2), v_2), \ldots, (k_S(v_L), v_L)\} \tag{2}$$

in which value $v_1 \in V$ appearing $k_S(v_1)$ times, value $v_2 \in V$ appearing $k_S(v_2)$ times and so on. So the value $k_S(v_i)$ for $i = 1,2,..., L$, specifies the number of occurrences of the value $v_i \in V$ in the multiset S, where $k_S(v_i) \geq 0$ (note that $v_i \in V$ is ordinary set notation). Let us assume that the values $v_i \in V$ for $k_S(v_i) = 0$ may be omitted and the multiset S can be described in a simplified form.

An ordinary set is a special case of a multiset. Any ordinary subset A from the set V can be identified with the multiset $\{(\chi_A(v), v)\}$, $\forall v \in V$ where $\chi_A(.)$ is its characteristic function, $\chi_A : V \to \{0,1\}$. Note, that $\chi_A(v) = 1$ if and only if $v \in A$, i.e., an multiset $S = \{(k_S(v), v)\}$ is an ordinary set if $k_S(v) = 0$ or 1, $\forall v \in V$. A multiset is a set if the multiplicity of every element is at most one.

A multiset S is empty, denoted by Ø or {}, if and only if $\forall i \in \{1,2,...,L\}$, $k_S(v_i) = 0$.

Example 1. Let us assume that S be the multiset drawn from a set V, where $V = \{a,b,c,d,e,f,g\}$. Exemplary multiset S can be described as $S = \{(3,a),(0,b),(0,c),(2,d),(0,e),(5,f),(0,g)\}$. The values $v_i \in V$ for $k_S(v_i) = 0$ may be omitted to simplify the notation and the multiset S can be described in a simplified form as $S = \{(3,a),(2,d),(5,f)\}$.

Basic notions of multiset are presented below.

The support or the *root* of the multiset S drawn from a set V, denoted by S^*, is an ordinary set defined as follows: $S^* = \{v \in V : k_S(v) > 0\}$ }. So, if $\forall v \in V$ such that $k_S(v) > 0$ this implies that $v \in S^*$, and $\forall v$ such that $k_S(v) = 0$ this implies that $v \notin S^*$. Note that the characteristic function of S^* can be described as $\chi_{S^*}(v) = \min\{k_S(v), 1\}$. Exemplary, the support of the multiset $S = \{(3,a),(2,d),(5,f)\}$ drawn from a set V can be described as $S^* = \{a,d,f\}$.

The cardinality of the multiset S, denoted by *card*(S) or $|S|$, is defined as the total number of its elements, i.e., $card(S) = \sum_{i=1}^{L} k_S(v_i)$.

The dimensionality of the multiset S, denoted by *dim*(S) or $/S/$, is defined as the total number of various elements, i.e., $dim(S) = \sum_{i=1}^{L} \chi_S(v_i)$. Exemplary, if $S = \{(3,a),(3,b)\}$ than $card(S) = 6$ and $dim(S) = 2$.

The multiset space $[V]^o$ is the set of all multisets whose elements are in V such that no element occurs more than o times, i.e., the multiset space can be described as $[V]^o = \{ \{(k_S(v_1),v_1), (k_S(v_2),v_2), ..., (k_S(v_L),v_L)\} : \textit{for } v_i \in V, i = 1,2,...,L \textit{ and } 0 \leq k_S(v_i) \leq o\}$.

The complement of the multiset S drawn from the set V in the multiset space $[V]^o$ is the multiset S^C such that $k_{S^C}(v) = o - k_S(v)$, $\forall v \in V$. Exemplary, if multiset $S = \{(5,a),(1,b)\}$ drawn from the set $V = \{a,b,c\}$ belongs to the multiset space $[V]^8$ than the complement of its can be described as $S^C = \{(3,a),(7,b),(8,c)\}$.

The upper cut of the multiset S, denoted by $\bar{S}$, is the ordinary set described as $\bar{S} = \{(k_{\bar{S}}(v_1),v_1),(k_{\bar{S}}(v_2),v_2), ...,(k_{\bar{S}}(v_L),v_L)\}$, where $k_{\bar{S}}(v_i) = \begin{cases} 1 & if \quad k_S(v_i) = 0 \\ 0 & if \quad k_S(v_i) \geq 1 \end{cases}$, $i \in \{1,2,...,L\}$.

The upper cut of the multiset S drawn from the set V is an ordinary set $\bar{S}$ with values don't contained in the multiset S only in the set V.

Corollary 1. *Assume that we have the multiset* S *drawn from the set* V. *The following property is satisfied:* $\bar{S} \oplus S^* = V$.

Example 2. Assume that exemplary multiset S drawn from the set V, where $V=\{a, b, c, d\}$, can be described as $S=\{(13,a),(12,d)\}$. Its support can be described as $S^* = \{a,d\}$ and the upper cut as $\bar{S} = \{b,c\}$. The addition of its support and the upper cut results in the entire set V, $S^* \oplus \bar{S} = \{a,d\} \oplus \{b,c\} = \{a, b, c, d\}$.

The concept of specificity provides a measure of the amount of information contained in a subset. Specificity measures for a fuzzy set were introduced by Yager (1982, 1990). The specificity is one (maximum value) only for crisp sets with just one element (singletons). The specificity measure of a set decreases when the number of its elements increases.

Let V be a finite set of nominal elements. Let consider the non-empty multiset S drawn from the set V, where $S \subseteq [V]^n$ and $card(S^*) \geq 1$. Here, we propose the following way to measure a level of quasi multiset's specificity.

Definition 2. *The measure of quasi specificity of the non-empty multiset* S, *normalized to the range 0-1, is defined in as follows*

$$MS(S) = \frac{card(V \setminus S^*)}{card(V) - 1} \tag{3}$$

Note, that measure of quasi specificity of multiset S is one (maximum value) only for the support S^* with just one element. Such sets may of course be plenty. The quasi specificity measure decreases when the number of its elements increases. Note that multiset S cannot be empty, by assumption. It is easy to notice that measure of quasi specificity of multiset satisfies the condition $0 \leq MS(S) \leq 1$. Note, that

1) $MS(S) = 1$ if and only if $card(S^*) = 1$,
2) if $S_1^* \subseteq S_2^*$, then $MS(S_1) \geq MS(S_2)$.

Example 3. Let us assume the set $V = \{a,b,c\}$. A few measures of the exemplary multisets S drawn from the set V belonging to the space $[V]^5$ are shown in Table 1.

Table 1. Measures S^C, S^*, $\bar{S}$ and $MS(S)$ of exemplary multiset S

S	The complement S^C in $[V]^5$	The support S^*	The upper cut $\bar{S}$	$MS(S)$
$\{(2,b)\}$	$\{(5,a),(3,b),(5,c)\}$	$\{b\}$	$\{a,c\}$	1
$\{(3,a),(5,b)\}$	$\{(2,a),(5,c)\}$	$\{a,b\}$	$\{c\}$	1/2
$\{(2,a),(1,b),(3,c)\}$	$\{(3,a),(4,b),(2,c)\}$	$\{a,b,c\}$	$\varnothing$	0

Let us assume that V is a finite set of nominal elements. Assume that S_1 and S_2 are two multisets drawn from the set V, Eq. (2). The rules of comparison of the multisets are presented below.

Equality. We say that two multisets S_1 and S_2 are *equal* or *the same*, denoted by $S_1 = S_2$, if $\forall v \in V$ the condition $k_{S_1}(v) = k_{S_2}(v)$ is satisfied. The following condition is fulfilled: if $S_1 = S_2$ than $S_1^* = S_2^*$, however the converse need not hold fulfilled.

Similarity. We say that two multisets S_1 and S_2 are *similar* if $\forall v \in V$, $v \in S_1$ if and only if $v \in S_2$. The similar multisets have equal support sets but need not be equal themselves. Exemplary, the multisets $S_1 = \{(3,a),(2,d),(5,f)\}$ and $S_2 = \{(5,a),(1,d),(3,f)\}$ are similar but not equal.

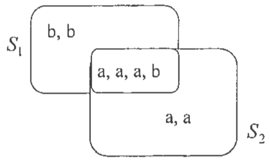
Inclusion. A multiset S_1 is *a sub-multiset* S_2, denoted as $S_1 \subseteq S_2$, if $\forall v \in V$ condition $k_{S_1}(v) \le k_{S_2}(v)$ is satisfied.

Many features of operations under multisets are analogues to features of operations under ordinary sets. The basic definitions and notions of the multisets are presented below. The following operations are defined on multisets (Table 2, Fig. 1):

Table 2.

Union $S_1 \cup S_2$	$S_1 \cup S_2 = \{(k_{S_1 \cup S_2}(v), v): \forall v \in V, k_{S_1 \cup S_2}(v) = \max\{k_{S_1}(v), k_{S_2}(v)\}\}$ If $S_1 = \{(3,a),(3,b)\}$ and $S_2 = \{(5,a),(1,b)\}$ than $S_1 \cup S_2 = \{(5,a),(3,b)\}$.
Intersection $S_1 \cap S_2$	$S_1 \cap S_2 = \{(k_{S_1 \cup S_2}(v), v): \forall v \in V, k_{S_1 \cap S_2}(v) = \min\{k_{S_1}(v), k_{S_2}(v)\}\}$ If $S_1 = \{(3,a),(3,b)\}$ and $S_2 = \{(5,a),(1,b)\}$ than $S_1 \cap S_2 = \{(3,a),(1,b)\}$.
Arithmetic addition $S_1 \oplus S_2$	$S_1 \oplus S_2 = \{(k_{S_1 \oplus S_2}(v), v): \forall v \in V, k_{S_1 \oplus S_2}(v) = k_{S_1}(v) + k_{S_2}(v)\}$ If $S_1 = \{(3,a),(3,b)\}$ and $S_2 = \{(5,a),(1,b)\}$ than $S_2 \oplus S_1 = \{(8,a),(4,b)\}$.
Arithmetic subtraction $S_1 \ominus S_2$	$S_1 \ominus S_2 = \{(k_{S_1 \ominus S_2}(v), v): \forall v \in V, k_{S_1 \ominus S_2}(v) = \max\{k_{S_1}(v) - k_{S_2}(v), 0\}\}$ or $S_1 \ominus S_2 = \{(k_{S_1 \ominus S_2}(v), v): \forall v \in V, k_{S_1 \ominus S_2}(v) = k_{S_1}(v) - k_{S_1 \cap S_2}(v)$ If $S_1 = \{(3,a),(3,b)\}$ and $S_2 = \{(5,a),(1,b)\}$ than $S_2 \ominus S_1 = \{(2,a),(0,b)\}$ and $S_1 \ominus S_2 = \{(0,a),(2,b)\}$.
Symmetric difference $S_1 \Delta S_2$	$S_1 \Delta S_2 = \{(k_{S_1 \Delta S_2}(v), v): \forall v \in V, k_{S_1 \Delta S_2}(v) = \lvert k_{S_1}(v) - k_{S_2}(v) \rvert\}$ If $S_1 = \{(3,a),(3,b)\}$ and $S_2 = \{(5,a),(1,b)\}$ than $S_1 \Delta S_2 = \{(2,a),(2,b)\}$.

$S_1 = \{(3,a),(3,b)\}$ $S_2 = \{(5,a),(1,b)\}$



$S_1 \cup S_2 = S_2 \cup S_1 = \{(5,a),(3,b)\}$

$S_1 \cap S_2 = S_2 \cap S_1 = \{(3,a),(1,b)\}$

$S_2 \oplus S_1 = S_1 \oplus S_2 = \{(8,a),(4,b)\}$

$S_2 \ominus S_1 = \{(2,a),(0,b)\}$, $S_1 \ominus S_2 = \{(0,a),(2,b)\}$

$S_1 \Delta S_2 = S_2 \Delta S_1 = \{(2,a),(2,b)\}$.

Fig. 1. An exemplary multisets S_1, S_2 and operations defined on multisets

Corollary 2. Let us assume that we have multisets drawn from the set V. The multisets operations $\oplus, \cup, \cap$ satisfies the following property

1. Commutativity:
   $S_1 \oplus S_2 = S_2 \oplus S_1$, $S_1 \cup S_2 = S_2 \cup S_1$, $S_1 \cap S_2 = S_2 \cap S_1$.
2. Associativity:
   $S_1 \oplus (S_2 \oplus S_3) = (S_1 \oplus S_2) \oplus S_3$,
   $S_1 \cup (S_2 \cup S_3) = (S_1 \cup S_2) \cup S_3$,
   $S_1 \cap (S_2 \cap S_3) = (S_1 \cap S_2) \cap S_3$.
3. Idempotence:
   $S_1 \cup S_1 = S_1$, $S_1 \cap S_1 = S_1$, $S_1 \oplus S_1 \ne S_1$.

4. Identity laws:

$S_1 \cup \emptyset = S_1$, $S_1 \cap \emptyset = \emptyset$, $S_1 \oplus \emptyset = S_1$.

5. Distributivity:

$S_1 \oplus (S_2 \cup S_3) = (S_1 \oplus S_2) \cup (S_1 \oplus S_3)$,

$S_1 \oplus (S_2 \cap S_3) = (S_1 \oplus S_2) \cap (S_1 \oplus S_3)$,

$S_1 \cup (S_2 \cap S_3) = (S_1 \cup S_2) \cap (S_1 \cup S_3)$,

$S_1 \cap (S_2 \cup S_3) = (S_1 \cap S_2) \cup (S_1 \cap S_3)$.

It is easy to see that operator $\oplus$ is stronger than both $\cup$ and $\cap$ in the sense that none of them distributes over $\oplus$, also $(S_1 \cap S_2) \subseteq (S_1 \cup S_2) \subseteq (S_1 \oplus S_2)$.

Assume that we have two multisets with finite support drawn from the set V. The following property is satisfied: $card(S_1 \cup S_2) + card(S_1 \cap S_2) = card(S_1) + card(S_2)$.

There are many existing methods for comparing the multisets in which the distances between multisets can be defined by different ways. Assume that we have two multisets S_1 and S_2 drawn from the set V, denoted by

$$S_1 = \{(k_{S_1}(v_1), v_1), (k_{S_1}(v_2), v_2), \ldots, (k_{S_1}(v_L), v_L)\}, \qquad (4)$$
$$S_2 = \{(k_{S_2}(v_1), v_1), (k_{S_2}(v_2), v_2), \ldots, (k_{S_2}(v_L), v_L)\}.$$

Three difference between two multisets introduced by Petrowsky (1994, 2001, 2003) for parameter p=1 and $w_i = 1$, $\forall i \in \{1, 2, \ldots, L\}$ are presented below.

$$d_{1,1}(S_1, S_2) = \sum_{i=1}^{L} \left| k_{S_1}(v_i) - k_{S_2}(v_i) \right|, \qquad (5)$$

$$d_{2,1}(S_1, S_2) = \frac{S_1 \Delta S_2}{S_1 \oplus S_2} \qquad (6)$$

$$d_{3,1}(S_1, S_2) = \frac{S_1 \Delta S_2}{S_1 \cup S_2} \qquad (7)$$

The distance $d_{2,1}(S_1, S_2)$ and $d_{3,1}(S_1, S_2)$ satisfy the normalization condition $0 \le d_{j,1}(S_1, S_2) \le 1$, for j=2,3. The distance $d_{3,1}(S_1, S_2)$ is not defined for $S_1 = S_2 = \emptyset$, so $d_{3,1}(\emptyset, \emptyset) = 0$ by the definition.

3. Measure of perturbation of the multisets

Let us assume that we have two multisets S_1 and S_2 drawn from the set V, $V = \{v_1, v_2, \ldots, v_L\}$, and $S_1, S_2 \subseteq [V]^o$, denoted by

$$S_1 = \{(k_{S_1}(v_1), v_1), (k_{S_1}(v_2), v_2), \ldots, (k_{S_1}(v_L), v_L)\},$$
$$S_2 = \{(k_{S_2}(v_1), v_1), (k_{S_2}(v_2), v_2), \ldots, (k_{S_2}(v_L), v_L)\}.$$

$\forall v_i \in V$, $i \in \{1, 2, \ldots, L\}$, $k_{S_1} : V \to \{0, 1, 2, \ldots, o\}$, $k_{S_2} : V \to \{0, 1, 2, \ldots, o\}$ where counting function $k_{S_1}(v_i)$ and $k_{S_2}(v_i)$ specify the number of occurrences of the element $v_i \in V$ in the multisets S_1 and S_2. Those elements which are not included in the multiset S have zero count.

The following condition $k_{S_1 \cap S_2}(v_i) + k_{S_1 \cup S_2}(v_i) = k_{S_1}(v_i) + k_{S_2}(v_i)$ is satisfied.

Attaching the first multiset S_1 to the second multiset S_2 can be considered that the second multiset is perturbed by the first multiset, in other words the multiset S_1 perturbs the multiset S_2 with some degree. The *result of perturbation the multiset S_2 by the multiset S_1* is $S_1 \Theta S_2$, denoted by $(S_1 \mapsto S_2)$.

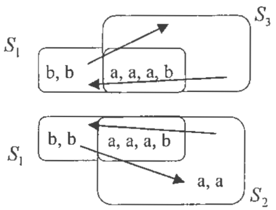
Example 4. Let us consider the set $V = \{a,b,c,d,e\}$ and an exemplary two multisets $S_1 = \{(1,a),(1,e)\}$ and $S_2 = \{(1,a),(1,d),(3,e)\}$, where $S_1, S_2 \subseteq [V]^4$. Multiset S_1 perturbs the multiset S_2 with the zero degree because a following condition is satisfied: $(S_1 \mapsto S_2) := S_1 \Theta S_2 = \emptyset$. On the other hand, multiset S_2 perturbs the multiset S_1 with the greater than zero degree because $(S_2 \mapsto S_1) := S_2 \Theta S_1 = \{(1,d),(2,e)\}$.

Here we propose the following measure of multiset's perturbation.

Definition 3. *Measure of perturbation of the multiset S_2 by the multiset S_1*, where $S_1, S_2 \subseteq [V]^n$, denoted by $Per(S_1 \mapsto S_2)$, is defined in the following manner:

$$Per(S_1 \mapsto S_2) = \frac{S_1 \Theta S_2}{S_1 \oplus S_2} = \frac{1}{L}\sum_{i=1}^{L} \frac{k_{S_1}(v_i) - k_{S_1 \cap S_2}(v_i)}{k_{S_1}(v_i) + k_{S_2}(v_i)} \tag{8}$$

Example 5. Let us consider the set $V = \{a,b\}$. An exemplary multisets $S_1, S_2, S_3 \subseteq [V]^5$ are shown in Fig.1, directions of perturbation are indicated by arrows.



$S_1 = \{(3,a),(3,b)\}$, $S_2 = \{(5,a),(1,b)\}$, $S_3 = \{(3,a),(1,b)\}$

$Per(S_1 \mapsto S_3) = \frac{1}{4}$, $Per(S_3 \mapsto S_1) = 0$

$Per(S_1 \mapsto S_2) = \frac{1}{4}$, $Per(S_2 \mapsto S_1) = \frac{1}{8}$

Fig. 1. An exemplary multisets $S_1, S_2, S_3 \subseteq [V]^5$

The measures of perturbation of the multisets $S_1, S_2, S_3 \subseteq [V]^5$ are show below

$$Per(S_1 \mapsto S_2) = \frac{1}{2}\sum_{i=1}^{2} \frac{k_{S_1}(v_i) - k_{S_1 \cap S_2}(v_i)}{k_{S_1}(v_i) + k_{S_2}(v_i)} = \frac{1}{2}\left(\frac{k_{S_1}(a) - k_{S_1 \cap S_2}(a)}{k_{S_1}(a) + k_{S_2}(a)} + \frac{k_{S_1}(b) - k_{S_1 \cap S_2}(b)}{k_{S_1}(b) + k_{S_2}(b)}\right) =$$
$$= \frac{1}{2}\left(\frac{3-3}{3+5} + \frac{3-1}{3+1}\right) = \frac{1}{4}$$

$$Per(S_2 \mapsto S_1) = \frac{1}{2}\sum_{i=1}^{2} \frac{k_{S_2}(v_i) - k_{S_1 \cap S_2}(v_i)}{k_{S_2}(v_i) + k_{S_1}(v_i)} = \frac{1}{2}\left(\frac{k_{S_2}(a) - k_{S_1 \cap S_2}(a)}{k_{S_2}(a) + k_{S_1}(a)} + \frac{k_{S_2}(b) - k_{S_1 \cap S_2}(b)}{k_{S_2}(b) + k_{S_1}(b)}\right) =$$
$$= \frac{1}{2}\left(\frac{5-3}{5+3} + \frac{1-1}{1+3}\right) = \frac{1}{8}$$

$$Per(S_1 \mapsto S_3) = \frac{1}{2}\sum_{i=1}^{2} \frac{k_{S_1}(v_i) - k_{S_1 \cap S_3}(v_i)}{k_{S_1}(v_i) + k_{S_3}(v_i)} = \frac{1}{2}\left(\frac{k_{S_1}(a) - k_{S_1 \cap S_3}(a)}{k_{S_1}(a) + k_{S_3}(a)} + \frac{k_{S_1}(b) - k_{S_1 \cap S_3}(b)}{k_{S_1}(b) + k_{S_3}(b)}\right) =$$

$$=\frac{1}{2}\left(\frac{3-3}{3+3}+\frac{3-1}{3+1}\right)=\frac{1}{4}$$

$$Per(S_3 \mapsto S_1)=\frac{1}{2}\sum_{i=1}^{2}\frac{k_{S_3}(v_i)-k_{S_1\cap S_3}(v_i)}{k_{S_1}(v_i)+k_{S_3}(v_i)}=\frac{1}{2}\left(\frac{k_{S_3}(a)-k_{S_1\cap S_3}(a)}{k_{S_1}(a)+k_{S_3}(a)}+\frac{k_{S_3}(b)-k_{S_1\cap S_3}(b)}{k_{S_1}(b)+k_{S_3}(b)}\right)=$$

$$=\frac{1}{2}\left(\frac{3-3}{3+3}+\frac{1-1}{3+1}\right)=0$$

Corollary 3. *Measure of perturbation of the multiset S_2 by the multiset S_1 satisfies the following condition* $0 \le Per(S_1 \mapsto S_2) \le 1$.

Proof. 1) We first prove the first inequality $Per(S_1 \mapsto S_2) \ge 0$. It should be noticed that the inequality $k_{S_1\cap S_2}(v_i) \le k_{S_1}(v_i)$, $\forall i \in \{1,2,...,L\}$ is satisfied, so $k_{S_1}(v_i)-k_{S_1\cap S_2}(v_i) \ge 0$. We obtain the following inequality

$$Per(S_1 \mapsto S_2)=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)-k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)} \ge 0 .$$

2) Let us prove now the second inequality, $Per(S_1 \mapsto S_2) \le 1$. It should be noticed that the inequality $k_{S_1\cap S_2}(v_i) \le k_{S_2}(v_i)$, $\forall i \in \{1,2,...,L\}$ is satisfied. We obtain the following inequality

$$Per(S_1 \mapsto S_2)=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)-k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)} \le \frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)+k_{S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}=1$$

Measure of perturbation of one set by another set satisfies the following properties:

Corollary 4. *The following condition is fulfilled* $Per(S_1 \mapsto S_2)=0$ *if and only if* $k_{S_1}(v_i)=k_{S_1\cap S_2}(v_i)$, $\forall i \in \{1,2,...,L\}$, i.e., $S_1 \subseteq S_2$

Corollary 5. *If the following condition is fulfilled* $k_{S_2}(v_i)=0$, $\forall i \in \{1,2,...,L\}$, i.e., $S_2=\varnothing$, *and* $k_{S_1}(v_i)>0$ *then condition* $Per(S_1 \mapsto S_2)=1$ *is satisfied.*

Proof. Suppose that that conditions $k_{S_2}(v_i)=0$, $\forall i \in \{1,2,...,L\}$ are satisfied. We obtain

$$Per(S_1 \mapsto S_2)=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)-k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)}{k_{S_1}(v_i)}=\frac{L}{L}=1.$$

Corollary 6. *If the following condition is fulfilled* $k_{S_1}(v_i)=k_{S_2}(v_i)$, $\forall i \in \{1,2,...,L\}$ *then condition* $Per(S_1 \mapsto S_2)=Per(S_2 \mapsto S_1)$ *is satisfied.*

Proof. Suppose that condition $k_{S_1}(v_i)=k_{S_2}(v_i)$, $\forall i \in \{1,2,...,L\}$ is satisfied. We obtain the following equation

$$Per(S_1 \mapsto S_2)=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)-k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_2}(v_i)-k_{S_1\cap S_2}(v_i)}{k_{S_2}(v_i)+k_{S_1}(v_i)}=Per(S_2 \mapsto S_1)$$

Corollary 7. *If the following condition is fulfilled* $k_{S_1\cap S_2}(v_i)=0$, $\forall i \in \{1,2,...,L\}$ *then condition* $Per(S_1 \mapsto S_2)+Per(S_2 \mapsto S_1)=1$ *is satisfied.*

Proof. Suppose that condition $k_{S_1\cap S_2}(v_i)=0$, $\forall i \in \{1,2,...,L\}$ is satisfied. We obtain the following equation

$$Per(S_1 \mapsto S_2)+Per(S_2 \mapsto S_1)=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)-k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}+\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_2}(v_i)-k_{S_1\cap S_2}(v_i)}{k_{S_2}(v_i)+k_{S_1}(v_i)}=$$

$$=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}+\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_2}(v_i)}{k_{S_2}(v_i)+k_{S_1}(v_i)}=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)+k_{S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}=1$$

Corollary 8. *If the following conditions are fulfilled* $k_{S_1\cap S_2}(v_i)\le k_{S_1\cap S_3}(v_i)$ *and* $k_{S_2}(v_i)\le k_{S_3}(v_i)$, $\forall i\in\{1,2,...,L\}$ *then the inequality* $Per(S_1\mapsto S_2)\ge Per(S_1\mapsto S_3)$ *is satisfied.*

Proof. Suppose that conditions $k_{S_1\cap S_2}(v_i)\le k_{S_1\cap S_3}(v_i)$ and $k_{S_2}(v_i)\le k_{S_3}(v_i)$, $\forall i\in\{1,2,...,L\}$ are satisfied. We obtain the following inequality

$$Per(S_1\mapsto S_2)=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)-k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}\ge\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)-k_{S_1\cap S_3}(v_i)}{k_{S_1}(v_i)+k_{S_3}(v_i)}=Per(S_1\mapsto S_3).$$

Corollary 9. *If the following conditions are fulfilled* $k_{S_1\cap S_2}(v_i)\le k_{S_1\cap S_3}(v_i)$ and $k_{S_2}(v_i)\le k_{S_3}(v_i)$, $\forall i\in\{1,2,...,L\}$ *then the inequality* $Per(S_1\mapsto S_2)+Per(S_2\mapsto S_3)\ge Per(S_1\mapsto S_3)$ *is satisfied.*

Proof. Suppose that conditions $k_{S_1\cap S_2}(v_i)\le k_{S_1\cap S_3}(v_i)$ and $k_{S_2}(v_i)\le k_{S_3}(v_i)$, $\forall i\in\{1,2,...,L\}$ are satisfied. By Corollary 8 the inequality $Per(S_1\mapsto S_2)\ge Per(S_1\mapsto S_3)$ is satisfied. By Definition 3 the inequality $Per(S_2\mapsto S_3)\ge 0$ is satisfied. So, the inequality $Per(S_1\mapsto S_2)+Per(S_2\mapsto S_3)\ge Per(S_1\mapsto S_3)$ is satisfied.

Corollary 10. *The following condition*

$$Per(S_1\mapsto S_2)+Per(S_2\mapsto S_1)=1-\frac{1}{L}\sum_{i=1}^{L}\frac{2\cdot k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)} \tag{9}$$

Is satisfied.

Proof. We obtain the following expression

$$Per(S_1\mapsto S_2)+Per(S_2\mapsto S_1)=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)-k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}+\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_2}(v_i)-k_{S_2\cap S_1}(v_i)}{k_{S_2}(v_i)+k_{S_1}(v_i)}=$$

$$=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)-k_{S_1\cap S_2}(v_i)+k_{S_2}(v_i)-k_{S_2\cap S_1}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}=\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)+k_{S_2}(v_i)-k_{S_1\cap S_2}(v_i)-k_{S_2\cap S_1}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}=$$

$$=\frac{1}{L}\sum_{i=1}^{L}\left(1-\frac{k_{S_1\cap S_2}(v_i)+k_{S_2\cap S_1}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}\right)=1-\frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1\cap S_2}(v_i)+k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}=1-\frac{1}{L}\sum_{i=1}^{L}\frac{2\cdot k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)}.$$

Definition 9. *Measure of similarity of the multisets* S_1 *and* S_2, *where* $S_1,S_2\subseteq[V]^n$, *denoted by* $Sim(S_1,S_2)$, *is defined in the following manner:*

$$Sim(S_1,S_2)=\frac{1}{L}\sum_{i=1}^{L}\frac{2\cdot k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)} \tag{10}$$

Corollary 11. *A measure of similarity is a function which assigns to every pair of the multisets a nonnegative number and satisfies the following equations:*

1. $Sim(S_1,S_1)=1$, by Definition 9,
2. $Sim(S_1,S_2)=0$ if and only if $k_{S_1\cap S_2}(v_i)=0$, $\forall i\in\{1,2,...,L\}$, i.e., $S_1\cap S_2=\emptyset$,
3. $Sim(S_1,S_2)=Sim(S_2,S_1)$, by Definition 9,

4. If $k_{S_1 \cap S_2}(v_i) \geq k_{S_1 \cap S_3}(v_i)$ and $k_{S_2}(v_i) \leq k_{S_3}(v_i)$ then $Sim(S_1,S_2)+Sim(S_2,S_3) \geq Sim(S_1,S_3)$.

Proof. Suppose that conditions $k_{S_1 \cap S_2}(v_i) \geq k_{S_1 \cap S_3}(v_i)$ and $k_{S_2}(v_i) \leq k_{S_3}(v_i)$, $\forall i \in \{1,2,...,L\}$ are satisfied. We obtain the following inequality

$$Sim(S_1,S_2)=\frac{1}{L}\sum_{i=1}^{L}\frac{2\cdot k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)} \geq \frac{1}{L}\sum_{i=1}^{L}\frac{2\cdot k_{S_1\cap S_3}(v_i)}{k_{S_1}(v_i)+k_{S_3}(v_i)} = Sim(S_1,S_3).$$

By Definition 9 the inequality $Sim(S_2,S_3) \geq 0$ is satisfied. So, the inequality $Sim(S_1,S_2)+Sim(S_2,S_3) \geq Sim(S_1,S_3)$ is satisfied.

5. The inequality $0 \leq Sim(S_1,S_2) \leq 1$ is satisfied.

Proof. It should be noticed that the inequality $k_{S_1\cap S_2}(v_i) \leq k_{S_1}(v_i)$ and $k_{S_1\cap S_2}(v_i) \leq k_{S_2}(v_i)$, $\forall i \in \{1,2,...,L\}$ are satisfied, so

$$Sim(S_1,S_2)=\frac{1}{L}\sum_{i=1}^{L}\frac{2\cdot k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)} = \frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1\cap S_2}(v_i)+k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)} \leq \frac{1}{L}\sum_{i=1}^{L}\frac{k_{S_1}(v_i)+k_{S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)} = 1.$$

It should be noticed that the inequality $k_{S_1\cap S_2}(v_i) \geq 0$, $\forall i \in \{1,2,...,L\}$ is satisfied, so

$$Sim(S_1,S_2)=\frac{1}{L}\sum_{i=1}^{L}\frac{2\cdot k_{S_1\cap S_2}(v_i)}{k_{S_1}(v_i)+k_{S_2}(v_i)} \geq 0$$

4. Multi attribute objects with nominal description – multiset approach

Let us consider a given finite set of objects $U=\{e_n\}$, indexed by n, $n=1,2,...,N$. The objects are described by K nominal attributes $A=\{a_1,...,a_K\}$ indexed by j, $j=1,...,K$. The set $V_{a_j}=\{v_{1,j},v_{2,j},...,v_{L_j,j}\}$ is the domain of the attribute $a_j \in A$, L_j denotes the number of nominal values of the attribute a_j, $L_j \geq 2$, $j=1,...,K$.

Let us assume that every non-empty subset of a finite set U is called a *group*. We assume that the description of a group g, $g \subseteq U$, are denoted by G_g. A number of different definitions of symbolic descriptions are available in the literature. In the paper, each non empty group g can be represented by an ordered collection of multisets drawn from the ordinary sets of values of nominal attributes $\{a_1,a_2,...,a_K\}$ describing objects, $\forall V_{a_j}$ for $j=1,2,...,K$.

Definition 10. *Every group of objects g, $g \subseteq U$, can be represented by an ordered collection of multisets drawn from the ordinary sets of values of the attributes describing objects belonging to this group,*

$$G_g = < S_{1,t(1,g)}, S_{2,t(2,g)}, ..., S_{K,t(K,g)} > \tag{11}$$

where the multiset $S_{j,t(j,g)} \subseteq [V_{a_j}]^N$, $1 \leq card(S_{j,t(j,g)}) \leq N$ for $j \in \{1,...,K\}$. Each j-th multiset $S_{j,t(j,g)}$ drawn from the ordinary set $V_{a_j}=\{v_{1,j},v_{2,j},...,v_{L_j,j}\}$ for the group of objects g, $g \subseteq U$, can be represented by a set of pairs

$$S_{j,t(j,g)}= \\ =\{(k_{S_{j,t(j,g)}}(v_{1,t(j,g)}),v_{1,t(j,g)}),\ (k_{S_{j,t(j,g)}}(v_{2,t(j,g)}),v_{2,t(j,g)}),\ ...,(k_{S_{j,t(j,g)}}(v_{L_j,t(j,g)}),v_{L_j,t(j,g)})\} \tag{12}$$

where $v_{i,t(j,g)} \in V_{a_j}$ for $j=1,...,K$, $i=1,2,...,L_j$. The value $k_{S_{j,t(j,g)}}(v_{i,t(j,g)})$ for $j=1,...,K$ specifies the number of occurrences of the value $v_{i,t(j,g)}$ in the multiset $S_{j,t(j,g)}$, where $k_{S_{j,t(j,g)}}(v_{i,t(j,g)}) \geq 0$. The index $i,t(j,g)$ for $j \in \{1,2,...,K\}$ and $i=1,2,...,L_j$, specifies which value of the attribute a_j is used in the object in the group g. This notation states that for the objects belonging to the group g, for the attribute a_j, the value $v_{1,t(j,g)}$ appearing $k_{S_{j,t(j,g)}}(v_{1,t(j,g)})$ times, the value $v_{2,t(j,g)}$ appearing $k_{S_{j,t(j,g)}}(v_{2,t(j,g)})$ times and so on. So the value $k_{S_{j,t(j,g)}}(v_{i,t(j,g)})$ for $i = 1,2,...,L_j$, specifies the number of occurrences of the value $v_{i,t(j,g)} \in V_{a_j}$ in the multiset $S_{j,t(j,g)}$.

For the group consisting of a single object, i.e., $g = \{e_1\}$, the multisets $S_{j,t(j,g)}$, for $j = 1,...,K$, defined in Eq. (12) simplifies to the form

$$S_{j,t(j,g)} = \{(1, v_{i(j),t(j,g)})\}, \tag{13}$$

where $v_{i(j),t(j,g)} \in V_{a_j}$, for $i(j) \in \{1,2,...,L_j\}$. This notation states that the attribute a_j, $j = 1, 2, \ldots, K$, can take values $v_{i(j),t(j,g)}$ for the object e_1. For the object e_1 and for the attribute a_j the value $v_{i(j),t(j,g)}$ appearing one time in the multiset $S_{j,t(j,g)}$, so $k_{S_{j,t(j,g)}}(v_{i(j),t(j,g)}) = 1$. Description of the group consisting one object simplifies to the form

$$< \{(1, v_{i(1),t(1,g)})\}, \{(1, v_{i(2),t(2,g)})\}, ..., \{(1, v_{i(K),t(K,g)})\} > \tag{14}$$

where $v_{i(j),t(j,g)} \in V_{a_j}$ for $j = 1,...,K$. This notation states that the attribute a_j takes the value $v_{i(j),t(j,g)}$ for the object e_1. The index $i(j),t(j,g)$ for $j \in \{1,2,...,K\}$ specifies which value of the attribute a_j is used in the object e_1, $i \in \{1,2,...,L_j\}$.

Equ. (14) can be treated as a generalization of representation of object described by K attributes by multiset.

Example 6. Let us consider the objects describe by three nominal attributes $A=\{a_1, a_2, a_3\}$, where the sets $V_{a_1}=\{a,b,c\}$, $V_{a_2}=\{d,e\}$ and $V_{a_3}=\{f,h,n\}$ are respectively the domain of this attributes. For three exemplary objects e_1 represented by $<\{(1,a)\},\{(1,d)\},\{(1,f)\}>$, e_2 represented by $<\{(1,c)\},\{(1,d)\},\{(1,n)\}>$ and e_3 represented by $<\{(1,a)\},\{(1,d)\},\{(1,n)\}>$, we can describe an exemplary group $g=\{e_1, e_2, e_3\}$ by an ordered collection of three multisets in the following form $G_g =< \{(2,a),(1,c)\},\ \{(3,d)\},\ \{(1,f),(2,n)\} >$.

We say, that an object e_n represented by $<\{(1, v_{i,t(1,e_n)})\}, \{(1, v_{i,t(2,e_n)})\}, ..., \{(1, v_{i,t(K,e_n)})\}>$, $v_{i,t(j,e_n)} \in V_{a_j}$ for $j = 1,...,K$, belongs to the group g represented by $G_g = < S_{1,t(1,g)}, S_{2,t(2,g)}, ..., S_{K,t(K,g)} >$, denoted by $e_n \in g$, if the following relations are satisfied:

$$\{v_{i,t(j,e_n)}\} \subseteq S_{j,t(j,g)},\quad \forall j \in \{1,...,K\},\ i \in \{1,2,...,L_j\} \tag{15}$$

Example 7. The exemplary object e_n represented by $<\{(1,c)\},\{(1,d)\},\{(1,i)\}>$, for $K = 3$, belongs to the group g_1 represented by $G_{g_1} =< \{(2,a),(1,c)\},\ \{(3,d)\},\ \{(1,i),(2,n)\} >$ and does not belong to the group g_2 represented by $G_{g_2} =< \{(2,a),(1,c)\},\ \{(3,d)\},\ \{(3,f)\} >$.

Let us consider a two groups g_1 and g_2 described as follows:

$G_{g_1} =< S_{1,t(1,g_1)}, S_{2,t(2,g_1)}, ..., S_{K,t(K,g_1)} >$ and
$G_{g_2} =< S_{1,t(1,g_2)}, S_{2,t(2,g_2)}, ..., S_{K,t(K,g_2)} >$,

where $S_{j,t(j,g_1)} \subseteq [V_{a_j}]^N$, $S_{j,t(j,g_2)} \subseteq [V_{a_j}]^N$, $j \in \{1,2,...K\}$. The group g_1 containing the objects $\{e_n :\ n \in J_{g_1} \subseteq \{1,...,N\}\}$, and the group g_2 containing the objects $\{e_n :\ n \in J_{g_2} \subseteq \{1,...,N\}\}$, where $J_{g_1} \cap J_{g_2} = \varnothing$.

Definition 11. *The join between the group g_1 and g_2 is a new group g_3 described as:*

$$G_{g_3} := G_{g_1} \oplus G_{g_2} = < S_{1,t(1,g_1)} \oplus S_{1,t(1,g_2)},\ S_{2,t(2,g_1)} \oplus S_{2,t(2,g_2)},\ ...,S_{K,t(K,g_1)} \oplus S_{K,t(K,g_2)} > . \quad (16)$$

contains objects $\{e_n :\ n \in J_{g_1} \cup J_{g_2}\}$.

The dominance of groups can be determined on the ground of the set theory. We say, that *description of group G_{g_1} dominates description of group G_{g_2}* (denoted by $G_{g_1} \succeq G_{g_2}$) if the clauses $S_{j,t(j,g_1)} \supseteq S_{j,t(j,g_2)}$, $\forall j,\ j = 1,...,K$, are satisfied. It should be noticed that dominance is a transitive relation, and following conditions are satisfied:

$$\text{if } G_{g_1} \succeq G_{g_2} \text{ and } G_{g_2} \succeq G_{g_3} \text{ then } G_{g_1} \succeq G_{g_3} . \quad (17)$$

Example 8. An exemplary group $G_{g_1} = < \{(2,a),(1,c)\},\ \{(3,d)\},\ \{(1,i),(2,n)\} >$ dominates group $G_{g_2} = < \{(1,c)\},\ \{(2,d)\},\ \{(2,n)\} >$ and does not dominates $G_{g_3} = < \{(2,a)\},\ \{(3,d)\},\ \{(3,h)\} >$.

Formula (8) can be applied for all multisets of description of groups, and in this case we consider attaching G_{g_1} to G_{g_2} , or in other words a perturbation of G_{g_2} by G_{g_1} . In this way we can introduce a definition of the measure of perturbation of one group by another.

Definition 12. *Measure of perturbation of G_{g_2} by G_{g_1} , denoted $Per(G_{g_1} \mapsto G_{g_2})$, is defined in the following manner:*

$$Per(G_{g_1} \mapsto G_{g_2}) = \frac{1}{K} \sum_{j=1}^{K} Per(S_{j,t(j,g_1)} \mapsto S_{j,t(j,g_2)}) . \quad (18)$$

It is easy to notice that (18) holds for non-empty multisets $S_{j,t(j,g_1)}, S_{j,t(j,g_2)}$, $\forall j,\ j = 1,...,K$, and normalized to interval 0 and 1, can be rewritten as follows

$$Per(G_{g_1} \mapsto G_{g_2}) = \frac{1}{K} \sum_{j=1}^{K} \left(\frac{1}{L_j} \sum_{i=1}^{L_j} \frac{k_{S_{j,t(j,g_1)}}(v_i) - k_{S_{j,t(j,g_1)} \cap S_{j,t(j,g_2)}}(v_i)}{k_{S_{j,t(j,g_1)}}(v_i) + k_{S_{j,t(j,g_2)}}(v_i) - 1} \right) \quad (19)$$

Measure of perturbation of G_{g_2} by G_{g_1} is zero if and only if G_{g_2} dominates G_{g_1} , which can be stated as a following corollary.

Corollary 12. *$Per(G_{g_1} \mapsto G_{g_2}) = 0$ if and only if $G_{g_2} \succeq G_{g_1}$.*

Additionally we can prove that a measure of the group's perturbation is always positive and less than 1, as shown in the Corollary 13.

Corollary 13. *Measure of perturbation of G_{g_2} by G_{g_1} satisfies the following inequality*

$0 \le Per(G_{g_1} \mapsto G_{g_2}) \le 1$.

5. Conclusions

In this paper we recalled the concept of multisets as well as several related definitions. The purpose is following, namely we want to define and propose the new measure of remoteness between multisets of nominal values. The conception is based on multiset-theoretic operations and the previous papers of the authors where the idea of perturbation of sets and measure of perturbation was introduced and discussed (Krawczak and Szkatuła, 2013a,b, 2014). Instead of considering distance between two multisets, we introduced a new idea of *perturbation one multiset by another,* and next we define a *measure of perturbation* of one multiset by another multiset.

Next, we defined a description of a group of objects as a K-tuple of multisets (an ordered collection of multisets. Instead of considering dissimilarities between groups, we introduced a *measure of perturbation of one group by another group.* This measure defines changes of one group after adding a second group, and vice versa. The idea of the measure of group's perturbation is based on a relation between two attributes' values represented as multisets.

In result we obtain an extended idea of similarities of two sets, and extension of similarities of two multisets, and at the end similarities of groups of multisets.

REFERENCES

El-Sayed A. Abo-Tabl (2013). Topological approximations of multisets. Journal of the Egyptian Mathematical Society, 21, pp. 123-132.

Girish K.P., Sunil Jacob John (2012). Multiset topologies induced by multiset relations. Information Sciences 188, pp. 298-313.

Krawczak M., Szkatuła G. (2013a). A new measure of groups perturbation. Proceedings of the 2013 Joint IFSA World Congress NAFIPS Annual Meeting, Edmonton, Canada, 2013, pp. 1291-1296.

Krawczak M., Szkatuła G. (2013b). On perturbation measure of clusters – application. ICAISC 2013, Lecture Notes in Artificial Intelligence, Vol. 7895, Part II, Springer, Berlin, 176-183.

Krawczak M., Szkatuła G. (2014). An approach to dimensionality reduction in time series. Information Sciences, 260, 15-36.

Santini S., Jain R. (1996) Similarity Queries in Image Databases, Proceedings of IEEE Conference on Computer vision and Pattern recognition.

Singh D., Ibrahim A.M., Yohanna T., Singh J.N. (2008). A systematization of fundamentals of multisets. Lecturas Matematicas, Vol. 29, pp. 33-48.

Singh D., Ibrahim A.M., Yohanna T., Singh J.N. (2007). An overview of the applications of multisets. Novi Sad J. Math. Vol. 37, No. 2, pp. 73-92.

Syropoulos A. (2001). Mathematics of multisets. In. C.S. Calude at al. (Eds.) Multiset Processing, LNCS 2235, pp. 347-358.

R.R. Yager (1982) Measuring tranquility and anxiety in decision making: an application of fuzzy sets. International Journal of General Systems 8, 139–146.

R.R. Yager (1990) Ordinal measures of specificity. International Journal of General Systems, 17, 57–72.