International Journal of Uncertainty,
Fuzziness and Knowledge-Based Systems
Vol. 17, Suppl. 1 (August 2009) 153–172

World Scientific
www.worldscientific.com

COMPARISON OF SETS AND MULTISETS

AXEL HALLEZ*, ANTOON BRONSELAER† and GUY DE TRÉ‡

Department of Telecommunications and Information Processing,
Ghent University, Sint-Pietersnieuwstraat 41, Ghent, B9000, Belgium
* *axel.hallez@ugent.be*
† *antoon.bronselaer@ugent.be*
‡ *guy.detre@ugent.be*

Received 19 September 2008
Revised 24 January 2009
Accepted 23 April 2009

The comparison of sets of objects is a research topic with applications in diverse fields such as computer science, biology and psychology. Since the introduction of the Jaccard index, many techniques have been proposed. This paper aims at extending an existing framework of comparison indices for sets. Firstly, the novel indices account for similarities between elements, rather than identity of elements as is the case for existing techniques. As a result, a richer framework of comparison indices is obtained. The use of fuzzy quantifiers in this framework is shown. Secondly, the machinery for sets is extended to the case of multisets, which results in two classes of comparison indices. The first class considers each element instance as a separate element, while the second class considers groups of elements instances as an atomic entity. The number of instances is then a property of this group, that is taken into account when calculating similarity between element groups.

Keywords: Sets; multisets; comparison indices.

1. Introduction

In the field of computer science, comparison of data is an often occurring task. In the most general form, two pieces of data are given as input of a comparison process and the output quantifies a predefined relation (e.g.: similarity, preference, satisfaction, ...) between the given data. Such comparison tasks are encountered in, among others, data fusion, data filtering, querying, entity identification and case based reasoning. In all these examples, the data to be compared have a specific structure. Although many examples of comparing unstructured or semi-structured data exist, the focus of this paper is on structured data (e.g. databases, OO-applications, ...). Many solutions to tackle comparison of structured data have been presented in the past decades.[1–4] In each of these models, a complex data structure is assumed. Instances of such complex structures are denoted by the term *object* here and the problem of comparing objects is denoted object matching. An object has the property that it can be structurally decomposed into sub-objects. For example, relational tables can be decomposed into attributes, OO-objects can be decomposed into members, which in turn can be objects. It follows that the problem of comparing objects is decomposed to new comparison problems. These new problems have the advantage that they are comparison problems in well known domains of simple data types. Examples of such data types are numerical values, strings and booleans. This paper deals with comparison of the data types *sets* and *multisets*. These data types are of interest as many valued data types often occur, especially in OO-environments. Further on, the techniques introduced in this paper are based on the framework of Dubois and Prade,[5] which is a framework for comparison of fuzzy sets. As a consequence, the introduced techniques are applicable on fuzzy sets and fuzzy multisets as well.

The comparison indices proposed in this paper differ from those presented in the unified framework of Dubois and Prade in two ways. Firstly, element similarities are taken into account. All existing indices are based on the same two-step process: (i) calculation of derived sets (e.g.: intersection, symmetrical difference, ...) and (ii) evaluation of these derived sets (e.g.: through fuzzy measures). An important observation is that in step (i), calculation of derived sets is based on element equality. In many cases of set comparison, set elements are considered as *object properties*. More specific, an object is modeled by summarizing the properties owned by the object. However, when the elements of sets are objects themselves, rather than properties, the use of equality becomes too stringent. For that purpose, indices defined in this paper, generalize element equality to *similarity*. Secondly, comparison of multisets is studied as an interesting generalization of the case of regular sets.

The remainder of the paper is structured as follows. In Sec. 2, the framework by Dubois and Prade for comparison of fuzzy sets is explained. Based on this framework, generalized comparison indices for (fuzzy) sets, taking element similarity into account, are defined in Sec. 3. Next, the use of these indices in a framework for comparison of (fuzzy) multisets is studied in Sec. 4. Finally, the most important contributions of this paper are summarized and some concluding remarks are given in Sec. 5.

2. Related Work

In literature, a large variety of indices for comparison of sets exists. The first such index is due to Jaccard.[6] For two sets A and B, the Jaccard index is given by:

$$S_{Jac}(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

Many other indices such as symmetrical Tversky indices, Dice index, ... exist.[7,8] It can be shown that these indices can be classified in equivalence classes with respect to the order induced on the compared sets.[9,10] As such, it can be proven that Jaccard and symmetrical Tversky are equivalent. Going even further, some non-equivalent indices are proven to be quasi-equivalent, based on fuzzy order-equivalence.[11]

The techniques introduced in this paper are based on the work of Dubois and Prade.[5] They provide a framework for comparison of fuzzy sets that is based on the concept of scalar fuzzy set evaluators.

Definition 1. Given a universe U, and $\tilde{\wp}(U)$ the set of fuzzy sets on U, a scalar fuzzy set evaluator on U is a mapping g from $\tilde{\wp}(U)$ to $[0,1]$ such that:

$$\begin{aligned} &\bullet\ g(\emptyset) = 0 \\ &\bullet\ g(U) = 1 \\ &\bullet\ A \subseteq B \Rightarrow g(A) \leq g(B) \end{aligned} \tag{2}$$

A scalar fuzzy set evaluator g is called *universal* if and only if $g(A) = 1 \Leftrightarrow A = U$ and *existential* if and only if $g(A) = 0 \Leftrightarrow A = \emptyset$. Next to these definitions by Dubois and Prade, the novel concept of a uniform evaluator is introduced and defined in this paper. A scalar fuzzy set evaluator will be called *uniform* if the image of $g(A)$ depends solely on the cardinality of A. For a uniform evaluator g, there always exists a increasing function $g_c : [0, |U|] \rightarrow [0,1]$ such that $\forall A \in \tilde{\wp}(U) : g_c(|A|) = g(A)$. For example, the uniform probability measure P is defined as $P(A) = \frac{|A|}{|U|}$ and $P_c(u) = u/|U|$. If a scalar fuzzy set evaluator g is non-uniform, the image of g depends on elements that are contained in the set. The symmetrical difference of sets (Δ) is an important concept in the framework of Dubois and Prade. For crisp sets, Δ is defined as:

$$A \Delta B = (A \cup B) \cap (\bar{A} \cup \bar{B}) = (\bar{A} \cap B) \cup (A \cap \bar{B}) \tag{3}$$

with $\bar{X}$ the complementary set of X. The identity in Eq. (3) remains valid if operators for intersection, union and complement are implemented by resp. min, max and $1 - x$. Formally:

$$\begin{aligned} \mu_{A\Delta B}(u) &= \min(\max(\mu_A(u), \mu_B(u)), \max(1 - \mu_A(u), 1 - \mu_B(u))) \\ &= \max(\min(1 - \mu_A(u), \mu_B(u)), \min(\mu_A(u), 1 - \mu_B(u))) \end{aligned} \tag{4}$$

In case another t-norm/t-conorm pair is used, the identity does not hold anymore. Based on scalar fuzzy set evaluator, and the set operations for fuzzy sets, Dubois and Prade define three types of comparison indices: *inclusion indices*, *partial matching indices* and *similarity indices*.

2.1. *Inclusion indices*

As the name suggests, an inclusion index indicates the extent to which one set contains another set. Dubois and Prade provide three axiomatic requirements for an inclusion index I:

$$\begin{aligned} &\bullet\ I(A,B) = 1 \Leftrightarrow \bar{A} \cup B = U \\ &\bullet\ A \cap B = \emptyset \Rightarrow I(A,B) = 0 \\ &\bullet\ I(A,B) \text{depends on a scalar evaluation of } \bar{A} \cup B, \text{ namely } g(\bar{A} \cup B) \end{aligned} \tag{5}$$

Due to the first and third constraint, g must be a universal evaluator. If A and B have disjoint supports, then $\bar{A} \cup B = \bar{A}$, so $g(\bar{A} \cup B) \in [g(\bar{A}), 1]$. Scaling this interval results in an inclusion index with the unit interval as image:

$$I(A, B) = \frac{g(\bar{A} \cup B) - g(\bar{A})}{1 - g(\bar{A})} \tag{6}$$

In the context of object matching, an inclusion index can be used whenever it is sufficient that the set belonging to one object, is a part of the set belonging to another object. Such is the case with selecting job candidates. In that case, the required profile for a job is compared to candidate profiles, where each profile is a set of skills. When selecting candidates, it is important that a candidate has all skills required for the job. However, candidate skills that are not required for the job, are redundant. In other words, the set of candidate skills must contain the job skills.

2.2. *Partial matching indices*

A partial matching index is a symmetrical index that evaluates the intersection of two sets. Each partial matching index PM must satisfy four axiomatic requirements:

- $PM(A, B) = 0 \Leftrightarrow A \cup B = \emptyset$
- $A \subseteq \text{core}(B) \vee B \subseteq \text{core}(A) \Rightarrow PM(A, B) = 1$
- $PM(A, B) = PM(B, A)$
- $PM(A, B)$ depends on a scalar evaluation of $A \cap B$, namely $g(A \cap B)$

(7)

Due to the first and the fourth constraint, g must be an existential evaluator. A natural way of constructing partial matching index is:

$$PM(A, B) = \frac{g(A \cap B)}{f(A, B)} \tag{8}$$

with f a commutative $\tilde{\wp}(U)^2 \to [0, 1]$ mapping such that $f(A, B) \geq g(A \cap B)$. To satisfy the second axiomatic condition, $f(A, B) = g(A)$ if $A \subseteq \text{core}(B)$. An appropriate choice for $f(A, B)$ is $\min(|A|, |B|)$. Partial matching indices are useful in object matching if a minimal overlap between many valued attributes is required. For example, assume a database with electronical devices and a query for devices that are compatible. One of the requirements is then that devices can be plugged into each other, which can be done by using a partial matching index that compares the inputs and outputs of devices.

2.3. *Similarity indices*

While a partial matching index evaluates the intersection of two sets, a similarity index evaluates the symmetrical difference between two sets. In the framework of

Dubois and Prade, a similarity index S satisfies the following axiomatic constraints:

$$\begin{array}{l} \bullet\ S(A,B) = 1 \Leftrightarrow A\Delta B = \emptyset \\ \bullet\ \mathrm{supp}(A) \cap \mathrm{supp}(B) = \emptyset \Rightarrow S(A,B) = 0 \\ \bullet\ S(A,B) = S(B,A) \\ \bullet\ S(A,B) \text{ depends on } g(\overline{A\Delta B}) \text{ or on } g(\bar{A} \cap B) \text{ and } g(A \cap \bar{B}) \end{array} \tag{9}$$

Based on this set of axiomatic requirements, three types of similarity indices are defined. Indices of the first type are based on $g(\overline{A\Delta B})$. In this case, g must be a universal evaluator. For two sets with disjoint supports, $A\Delta B = A \cup B$, so $g(A \cup B)$ is a lower limit for $g(A\Delta B)$. Hence, after scaling, a similarity index of the first type, with image $[0, 1]$, is given by:

$$S(A,B) = \frac{g(\overline{A\Delta B}) - g(\overline{A \cup B})}{1 - g(\overline{A \cup B})} \tag{10}$$

Substitution of g with its derived existential evaluator g', such that $g'(A) = 1 - g(\bar{A})$, yields the following equivalent expression:

$$S(A,B) = \frac{g'(A \cup B) - g'(A\Delta B)}{g'(A \cup B)} \tag{11}$$

If g is the uniform probability measure then the Jaccard index is obtained:

$$S(A,B) = \frac{|A \cap B|}{|A \cup B|} = S_{Jac}(A,B) \tag{12}$$

Similarity indices of the second type are constructed by using a symmetrical function f of $g(\bar{A} \cap B)$ and $g(A \cap \bar{B})$. If Δ is defined based on $\cup$ then $A\Delta B = \emptyset \Rightarrow A \cup \bar{B} = \bar{A} \cup B = U$. It follows that $f(1,1)$ must equal 1. If $\mathrm{supp}(A) \cap \mathrm{supp}(B) = \emptyset$, then $\bar{A} \cup B = \bar{A}$ and $A \cup \bar{B} = \bar{B}$. Hence, a normalized similarity index of the second type is given by:

$$S(A,B) = \frac{f\left(g\left(A \cup \bar{B}\right), g\left(\bar{A} \cup B\right)\right) - f\left(g\left(\bar{A}\right), g\left(\bar{B}\right)\right)}{1 - f\left(g\left(\bar{A}\right), g\left(\bar{B}\right)\right)} \tag{13}$$

Finally, the third type of indices are based on a symmetrical combination of an inclusion index. Hence, this type of indices is indirectly based on $g(A \cap \bar{B})$ and $g(\bar{A} \cup B)$. They are given by:

$$S(A,B) = h\left(I\left(A,B\right), I\left(B,A\right)\right) \tag{14}$$

with h a commutative function satisfying $h(x,y) = 1 \Leftrightarrow x = y = 1$ and $h(0,0) = 0$.

2.4. *Related approaches*

Marín *et al.*[3] introduced an inclusion index and a derived similarity index for fuzzy sets that takes similarities of elements into acount. Let A and B be two fuzzy sets in a universe X and s a similarity relation defined over X. The inclusion of A into B, denoted $I_s(A,B)$, is defined as follows:

$$I_s(A,B) = \underset{x \in X}{min}\ \underset{y \in X}{max}\ \theta_{A,B,s}(x,y) \tag{15}$$

with:

$$\theta_{A,B,s}(x,y) = T(I(\mu_A(x), \mu_B(y)), s(x,y)) \tag{16}$$

where T is a t-norm and I an implicator.

Using this inclusion index, one can build a similarity index as follows:

$$S_s(A,B) = T(I_s(A,B), I_s(B,A)) \tag{17}$$

Because this similarity index can result in relatively high similarities, even if the cardinalities of the sets are very different, it can be necessary to correct the result by using the cardinality ratio $\frac{min(|A|,|B|)}{max(|A|,|B|)}$, provided that $max(|A|,|B|) > 0$.

3. Comparison Indices for Sets

In this section, the framework of comparison indices introduced in Sec. 2 is extended towards a framework that accounts for similarities between elements. As already mentioned, a classical setting assumes the universe U to be a list of properties an object can have. In such a setting, each element specifies a particular property and comparison of sets comes down to verification of shared properties. The use of element equality is sufficient in this case. However, when comparing many valued attributes in an object matching setting, the 'element-of' relation is not necessarily compatible with the 'is-a-property-of' relation. It is also possible that similarities between elements must be taken into account, for example because element equality is not informative enough in an object matching context. Moreover, similarity between two sets should measure similarities between elements, rather than measuring the extent to which sets contain the same elements. As an example, consider a social network site where people can maintain a profile. It is then of interest to search for people with similar interests. Assume that a profile contains a 'hobby' field which is clearly a many valued attribute. When searching for similar profiles, similarity between hobbies should not only search for people with the same hobbies, but also people with similar hobbies.

Justified by these considerations, it is studied in the following how a comparison index on the domain $\wp(U)$ (or $\tilde{\wp}(U)$ in a more general case) can be constructed, thereby using a comparison index on the domain U. For the determination of similarity between elements, it is assumed that a similarity measure s on the universe U is given. An important aspect is that all comparison indices are based on derived sets (intersection, union, difference). Obtaining these sets is based on element identity. Further on, scalar fuzzy set evaluators can also be based on element identity as non-uniform evaluators depend on the exact elements. Replacing element identity with element similarity in all these calculations is a non trivial problem. In many cases, similarity measures are assumed to be strictly reflexive, which means that $\forall(u,v) \in U^2 : s(u,v) = 1 \Leftrightarrow u = v$. In what follows, unless explicitly stated otherwise, this assumption is rejected as in many applications of set comparison, elements of the sets are compared on a part of their properties. Hence, it is possible that elements have only equal values for these properties and are considered equivalent in the context of the comparison.

Before the three classes of comparison indices are discussed, attention is given to the extension of set operators. As an example, assume the intersection operator ($\cap$) is to be extended to an operator $\cap_s$, that is based on the similarity measure s. For two sets A and B, and elements $a \in A$ and $b \in B$, with $a \neq b$, but $s(a, b) > 0$, the question is which of these elements belongs in the intersection. Adding them both would increase the intersection cardinality with one, which should be avoided. However choosing one is fully arbitrary. An option could be to put both elements in the intersection such that:

$$\mu_{A\cap_s B}(a) + \mu_{A\cap_s B}(b) = 1 \text{ and } \frac{\mu_{A\cap_s B}(a)}{\mu_{A\cap_s B}(b)} = \frac{\mu_A(a)}{\mu_B(b)}. \tag{18}$$

The solution proposed here avoids the actual determination of the extended intersection. It is assumed that a comparison index uses a uniform scalar fuzzy set evaluator, which means that only the number of common elements is important. Hence, only the cardinality of the extended intersection is calculated, which avoids the selection problem as described above. Of course, the assumption of a uniform evaluator implies that specific elements weights can not be used. The proposed determination of the number of common elements of two (fuzzy) sets, is based on a one-to-one relation between the fuzzy two sets. Assume A and B, two (fuzzy) sets in the universe U, and s a similarity measure over U. Let R_s be a one-to-one fuzzy relation (i.e. each element of A and B can be an element of at most one couple of R_s), such that:

$$\forall (a, b) \in \text{supp}(R_s) : \mu_{R_s}(a, b) = T'\left(T\left(\mu_A\left(a\right), \mu_B\left(b\right)\right), s\left(a, b\right)\right) \tag{19}$$

with T and T' t-norms and such that the cardinality of R_s:

$$|R_s| = \sum_{(a,b)\in \text{supp}(R_s)} \mu_{R_s}(a, b) \tag{20}$$

is maximal. In general, any choice for T and T' is allowed. However, for $(a, b) \in R_s$, $\mu_{R_s}(a, b)$ is in many cases required to be proportional with $s(a, b)$. Such a proportional effect is obtained by choosing the probabilistic t-norm. Therefor, in the remainder of this paper, it is assumed that $T'(x, y) = xy$, without loss of generality. In that case, R_s reduces to:

$$\forall (a, b) \in \text{supp}(R_s) : \mu_{R_s}(a, b) = T\left(\mu_A\left(a\right), \mu_B\left(b\right)\right) s\left(a, b\right) \tag{21}$$

The summation in Eq. (20) is a measure for the number of common elements in both sets, taking into account s. Based on this measure, the comparison indices in the framework of Sec. 2 are extended. Note that R_s is not bound to be unique. In case more than one R_s exists, the R_s with minimal $|\text{supp}(R_s)|$ is chosen. The relation

R_s allows for construction of derived sets. It is possible to define the projections $p_1(R_s)$ and $p_2(R_s)$ of R_s as follows:

$$\begin{aligned} &\forall a \in U : \mu_{p_1(R_s)}(a) = \sup_{b\in U} \mu_{R_s}(a,b) \\ &\forall b \in U : \mu_{p_2(R_s)}(b) = \sup_{a\in U} \mu_{R_s}(a,b) \end{aligned} \tag{22}$$

It is clear that $p_1(R_s) \subseteq A$ and $p_2(R_s) \subseteq B$. The projections consist of elements of A, respectively B, that have a corresponding element in the other set under s. The membership degree of an element in the projection reflects both the similarity with the corresponding element and the importance of both elements in the sets they are contained in. Having these projections, the elements that have no corresponding element in the other set under s are given by respectively $A\backslash p_1(R_s)$ and $B\backslash p_2(R_s)$. The union of these sets $(A\backslash p_1(R_s) \cup B\backslash p_2(R_s))$ fulfills the role of the symmetrical difference (Δ). With the relation R_s and its derived sets at hand, the comparison indices from Sec. 2 can be redefined. As will become clear, this requires modification of the axiomatic constraints.

3.1. *Inclusion indices*

The three axiomatic constraints of an inclusion index I are given in Eq. (5). These constraints are reviewed in a generalized context. The first constraint states that inclusion is complete if and only if $\bar{A} \cup B = U$. If a similarity measure s is to be taken into account, this constraint can only be valid if s is strict reflexive ($s(a,b) = 1 \Leftrightarrow a = b$). The second constraint ($A \cap B = \emptyset \Rightarrow I(A,B) = 0$) means that the inclusion index must be 0 if the sets have no elements in common. In the generalized framework, $|R_s|$ is a measure for the number of the common elements. Hence, the second axiom translates to: $R_s = \emptyset \Rightarrow I_s(A,B) = 0$. The third constraint states that an inclusion index must depend on $g(\bar{A} \cup B)$. However, it is hard to incorporate s into $\bar{A} \cup B$ and thus, a complete generalization of inclusion indices is impossible. Nevertheless, it is shown how specific cases can be generalized. To do so, a connection is required between the measure of conclusion and the extent to which elements of A occur in B. This connection can be implemented in two ways. Firstly, in some cases (depending on the scalar evaluator g and the used set operations), it is possible to reduce Eq. (6) to an evaluation of $A \cap B$ to A:

$$I(A,B) = \frac{g(\bar{A} \cup B) - g(\bar{A})}{1 - g(\bar{A})} = \frac{g(A \cap B)}{g(A)} \tag{23}$$

These inclusion indices can be generalized as follows:

$$I_s(A,B) = \frac{g(p_1(R_s))}{g(A)} \tag{24}$$

This represents that the measure of inclusion is an evaluation of the elements of A that have a corresponding element in B under s against an evaluation of A itself. A second way of generalizing inclusion indices is the use of an evaluation of $\overline{A \backslash p_1(R_s)}$. Formally:

$$I_s(A, B) = \frac{g\left(\overline{A \backslash p_1(R_s)}\right) - g\left(\bar{A}\right)}{1 - g\left(\bar{A}\right)} \tag{25}$$

This second generalization is conceptually closer to the original definition of inclusion indices.

Example 1. Assume we have a universe $U = \{a, b, c, d, e, f\}$ and a similarity relation s defined over U. The similarities of the elements in U according to s are listed in Table 1. Now let us have a look at two fuzzy sets in U: $A = \{a_{/1}, b_{/.6}, c_{/.2}, d_{/1}, e_{/1}\}$ and $B = \{b_{/1}, d_{/1}, f_{/.4}\}$. In order to calculate the inclusion degree of A into B or vice versa, we first need to find the optimal mapping R_s for A and B. Using the minimum t-norm to combine the membership degrees and the product t-norm to apply the similarities, we find that $R_s = \{(a, b)_{/1}, (d, d)_{/1}, (e, f)_{/.16}\}$ is the most optimal A to B mapping with a cardinality of 2.16.

If we chose the uniform probability measure for g, we can apply Eq. (24) to calculate the inclusion of A into B:

$$I_s(A, B) = \frac{g(p_1(R_s))}{g(A)} = \frac{|p_1(R_s)|}{|A|} = \frac{2.16}{3.8} \approx 0.57 \tag{26}$$

The inclusion of B into A gives us this result:

$$I_s(B, A) = \frac{g(p_2(R_s))}{g(B)} = \frac{|p_2(R_s)|}{|B|} = \frac{2.16}{2.4} = 0.9 \tag{27}$$

This inclusion index reflects the ratio of the number of elements in the first set that are mapped to a similar element by R_s to the total amount of elements in this set.

If we choose the infimum ($g(Z) = inf_{x \in U}\, \mu_Z(x)$) for the set evaluator, we need to use Eq. (25). We find following results (considering that $inf_{x \in U}\, \bar{Z} = 0$ for every

Table 1. Similaties in the example universe.

	a	b	c	d	e	f
a	1	1	.8	0	0	.4
b	1	1	.5	0	0	.2
c	.8	.5	1	.2	.4	.7
d	0	0	.2	1	1	0
e	0	0	.4	1	1	.4
f	.4	.2	.7	0	.4	1

normalized set Z):

$$I_s(A,B) = \frac{inf_{x\in U}\,\overline{A\backslash p_1(R_s)} - inf_{x\in U}\,\overline{A}}{1 - inf_{x\in U}\,\overline{A}} = \inf_{x\in U} \overline{A\backslash p_1(R_s)} \tag{28}$$

$$= \inf_{x\in U} \overline{\{b_{/.6}, c_{/.2}, e_{/.84}\}} = 0.16 \tag{29}$$

$$I_s(B,A) = \frac{inf_{x\in U}\,\overline{B\backslash p_2(R_s)} - inf_{x\in U}\,\overline{B}}{1 - inf_{x\in U}\,\overline{B}} = \inf_{x\in U} \overline{B\backslash p_2(R_s)} \tag{30}$$

$$= \inf_{x\in U} \overline{\{e_{/.24}\}} = 0.76 \tag{31}$$

It is clear that this inclusion index reflects the degree to which all of the elements of the first set are represented in the second set.

3.2. *Partial matching indices*

A partial matching index is described in Sec. 2 as an evaluation of the intersection of two sets. As already mentioned, a generalization of the intersection that takes into account similarities between elements is not provided here. Instead, the cardinality of R_s represents an indication of the number of common elements. Hence, it is possible to make a generalization of partial matching indices. Due to the use of R_s, the scalar fuzzy set evaluator is bound to be uniform.

As with generalized inclusion indices, the axiomatic constraints of partial matching indices (see Eq. (7)) are reviewed. The first constraint, stating that there is no partial match if the intersection is empty and vice versa, needs redefinition. In a generalized context, there is no partial matching if there is no similarity between elements. Hence, the constraint translates to $PM_s(A,B) = 0 \Leftrightarrow R_s = \emptyset$. The second and third constraint can be preserved. The fourth constraint, stating that the partial matching index depends on $g(A \cap B)$ is replaced by the constraint that the partial matching index depends on a scalar evaluation of $|R_s|$. As g must be a uniform evaluator, there exists a function g_c for which $g_c(|A|) = g(A)$. Consequently, Eq. (8) can be rewritten as:

$$PM(A,B) = \frac{g_c(|A \cap B)|}{f(A,B)} \tag{32}$$

By replacing $|A \cap B|$ with $|R_s|$, a generalized partial matching index is obtained:

$$PM_s(A,B) = \frac{g_c(|R_s|)}{f(A,B)} \tag{33}$$

An alternative for this generalization is an evaluation of $p_1(R_s)$ and $p_2(R_s)$ as follows:

$$PM_s(A,B) = f\left(\frac{g\,(p_1(R_s))}{g\,(A)}, \frac{g\,(p_2(R_s))}{g\,(B)}\right). \tag{34}$$

In that case, f is a commutative $[0,1]^2 \to [0,1]$ mapping.

Example 2. Using the same setting as in Example 1, we can calculate the partial matching degree of sets A and B.

Using the uniform probability measure for g and min of cardinalities for f we derive this result:

$$PM_s(A, B) = \frac{g_c(|R_s|)}{f(A, B)} = \frac{2.16}{2.4} = 0.9 \tag{35}$$

Note that this result is equal to $max(I_s(A, B), I_s(B, A))$ if we use the same set evaluator for the inclusion.

Using the supremum $(g(Z) = sup_{x \in U}\, \mu_Z(x))$ we need to apply Eq. (34). We use max of cardinalities for f. Taking into acount that $sup_{x\in U}\, \mu_Z = 1$ for any normalized set Z, we obtain following result:

$$PM_s(A, B) = max\left(\frac{sup_{x\in U}\, \mu_{p_1(R_s)}(x)}{sup_{x\in U}\, \mu_A(x)}, \frac{sup_{x\in U}\, \mu_{p_2(R_s)}(x)}{sup_{x\in U}\, \mu_B(x)}\right) = \underset{x\in U}{sup}\ \ \mu_{R_s}(x) = 1 \tag{36}$$

3.3. *Similarity indices*

The similarity of two sets is determined by the differences between two sets, which is given by $A\Delta B$. If element similarities are to be taken into account, $A \backslash p_1(R_s) \cup B\backslash p_2(R_s)$, can be used to model the differences between sets. The definition of a similarity index for sets is extended in the following way:

Definition 2. Let U be a universe and s a similarity measure over U. A similarity index for $(\tilde{\wp}(U), s)$ is a relation S_s that satisfies three properties:

$$\begin{aligned} &\bullet\ S_s(A, B) = 1 \Leftrightarrow A\backslash p_1(R_s) \cup B\backslash p_2(R_s) = \emptyset \\ &\bullet\ R_s = \emptyset \Leftrightarrow S_s(A, B) = 0 \\ &\bullet\ S_s(A, B) = S_s(B, A) \end{aligned} \tag{37}$$

The direct connection between $A\backslash p_1(R_s) \cup B\backslash p_2(R_s)$ and the actual similarity is hard to determine. In the preceding, the union of the two sets is used to normalize the evaluation of the difference. Using this technique in the generalized context implies a skew ratio, because on the one hand, equal elements are taken into account one time in the union. Similar but non-equal elements on the other hand, are both part of the union. In an extreme case, it is possible that two sets are completely equivalent with respect to similarity of elements ($|R_s| = |A| = |B|$ and $A\backslash p_1(R_s) \cup B\backslash p_2(R_s) = \emptyset$), but $A \cap B = \emptyset$. When using a uniform evaluator, this problem can be avoided. The cardinality of the union of A and B is equal to $|A| + |B| - |A \cap B|$. For fuzzy sets, this identity is only valid if $\mu_A(u) + \mu_B(u) = \mu_{A\cup B}(u) + \mu_{A\cap B}(u)$, which is the case for some of the most frequently used combinations of t-norm and t-conorm: $T_{\min}$ and $S_{\max}$, T_P and S_P (probabilistic), T_L and S_L (Lucasiewics). It follows that, in case of a uniform evaluator (again using the function g_c) and based on Eq. (11), a similarity index can be reduced to:

$$S(A, B) = \frac{g_c(|A \cap B|)}{g_c(|A \cup B|)} \tag{38}$$

In this expression, the cardinality of the intersection can be replaced by the number of common elements $|R_s|$ and the cardinality of the union by the sum of the cardinalities of the sets, minus the number of common elements:

$$S_s(A,B) = \frac{g_c(|R_s|)}{g_c(|A|+|B|-|R_s|)} \tag{39}$$

Example 3. Again using the same setting as in Example 1, we can calculate the similarity degree of sets A and B using a uniform probability measure as a set evaluator:

$$S_s(A,B) = \frac{g_c(|R_s|)}{g_c(|A|+|B|-|R_s|)} = \frac{2.16}{3.8+2.4-2.16} = \frac{2.16}{4.04} \approx 0.53 \tag{40}$$

3.4. *Modification with fuzzy quantifiers*

The use of scalar fuzzy set evaluators as described in the preceding, yields flexibility to some extent. A problem with such evaluators is the impossibility to ensure that an equal ratio of cardinalities in nominator and denominator, always yields the same result (unless for $g_c(u) = u/|U|$). This is caused by independent evaluation in nominator and denominator. For that purpose, an alternative method for calculation of the indices is proposed by using fuzzy quantifiers.[12] Fuzzy quantifiers are an extension of classical quantifiers 'for all' ($\forall$) and 'there exists' ($\exists$), which stem from predicate logic. Their purpose is to model less exact quantitative concepts like 'some', 'a few', 'almost all', ... In general, two types of fuzzy quantifiers are distinguished: absolute quantifiers and relative quantifiers. Absolute quantifiers are only dependent on the number of elements in a set (i.e. the number of entities that satisfy a proposition). Relative quantifiers are dependent on the ratio of elements that satisfy a proposition over the number of elements in the universe. The distinction between the two types leads to a different mathematical modeling. An absolute quantifier is modeled as a $\mathbb{R} \to [0,1]$ mapping that is applied directly on the cardinality of a set. A relative quantifier is a $[0,1] \to [0,1]$ mapping that is scaled with the size of the universe on which the quantifier is applied. Let Q be an absolute quantifier, then the similarity of two sets can be expressed by:

$$S_{Q,s}(A,B) = Q(|R_s|) \tag{41}$$

It is clear that absolute quantifiers are not very suitable to express the similarity of two sets, because in general, it is required to measure the number of common elements with respect to the total number of elements in both sets. Moreover, use of an absolute quantifier implies that the first constraint for generalized similarity indices (see Definition 2) can not be satisfied. For the definition of inclusion indices based on absolute quantifiers, the same difficulties are present, but for partial matching, absolute quantifiers a extremely useful. For a partial overlap it is sufficient that only a part of both sets is commonal. It is not required that this number of common elements is in proportion to the total number of elements in one of both sets or in the union. With use of an absolute quantifer Q, the following partial matching index is obtained:

$$PM_{Q,s}(A,B) = Q(|R_s|). \tag{42}$$

Hereby, Q expresses the extent to which the number of common elements in both sets is sufficient to have a partial match. In order to be a correct indicator, it is required that Q is a monotonic increasing function.

Unlike absolute quantifiers, relative quantifiers are suitable to define similarity indices for sets. Assume a relative quantifier Q, then the similarity of two sets is expressed by:

$$S_{Q,s}(A,B) = Q\left(\frac{|R_s|}{|A| + |B| - |R_s|}\right) \tag{43}$$

Hereby, the fraction expresses the ratio of the number of common elements over the number of different elements, taking into account the similarity measure s. Quantifier Q evaluates the extent to which this ratio is acceptable. A similarity index defined as in Eq. (43) satisfies Definition 2 only if Q is a monotonic increasing function with $x = 0 \Leftrightarrow Q(x) = 0$ and $x = 1 \Leftrightarrow Q(x) = 1$. Relative quantifiers are also suitable to define inclusion indices. With Q a relative quantifier, an inclusion index can defined by:

$$I_{Q,s}(A,B) = Q\left(\frac{|R_s|}{|A|}\right) \tag{44}$$

This index gives the extent to which the ratio of elements from A with a corresponding element in B over the total number of elements in A is satisfiable. Finally, a relative quantifier Q also allows for definition of a partial matching index:

$$PM_{Q,s}(A,B) = Q\left(\frac{|R_s|}{f(|A|,|B|)}\right). \tag{45}$$

Hereby, f is a symmetrical monotonic increasing function. In summarization of this section, the existing framework of comparison indices (Sec. 2) has been generalized to take element similarities into account. It has also been shown how fuzzy quantifiers can be used to obtain comparison indices of the generalized kind.

Example 4. For this example we will use two fuzzy quantifiers Q_1 which is an absolute quantifier and Q_2 which is a relative quantifier. Their membership functions are shown in Figure 1. Assume the same setting as in Example 1. Using an absolute quantifier Q_1 to calculate the partial matching degree of sets A and B yields:

$$PM_{Q_1,s}(A,B) = Q(|R_s|) = Q(2.16) = 0.58 \tag{46}$$

The partial matching degree of sets A and B can also be calculated by using a relative quantifier Q_2. Using min for f yields:

$$PM_{Q_2,s}(A,B) = Q\left(\frac{|R_s|}{f(|A|,|B|)}\right) = Q\left(\frac{2.16}{2.4}\right) = 1 \tag{47}$$

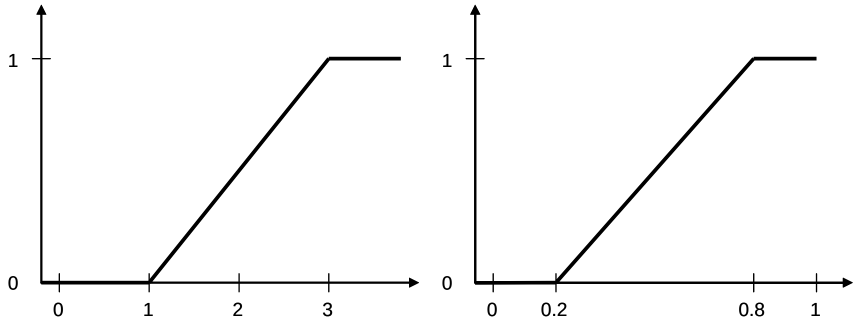

Fig. 1. Quantifiers Q_1 and Q_2.

Use of a relative quantifier also allows for the calculation of inclusion degrees:

$$I_{Q_2,s}(A,B) = Q_2\left(\frac{|R_s|}{|A|}\right) = Q_2\left(\frac{2.16}{3.8}\right) \approx 0.61 \tag{48}$$

$$I_{Q_2,s}(B,A) = Q_2\left(\frac{|R_s|}{|B|}\right) = Q_2\left(\frac{2.16}{2.4}\right) = 1 \tag{49}$$

and a similarity degree:

$$S_{Q_2,s}(A,B) = Q_2\left(\frac{|R_s|}{|A|+|B|-|R_s|}\right) = Q_2\left(\frac{2.16}{2.4+3.8-2.16}\right) \approx 0.56 \tag{50}$$

4. Comparison Indices for Multisets

4.1. *Multisets*

The generalized comparison indices defined in the previous section are indices for (fuzzy) sets. In this section it is studied how these ideas can be extended for (fuzzy) multisets. Multisets are a generalization of regular sets as a specific element can occur more than once. The concept of a multiset is defined as a fundamental entity by Yager[13] and a formal theory is developed by Blizzard.[14,15] Blizard defined multisets as first class entities rather than defining them within the framework of regular sets. Instead, it is shown by Blizard from a theoretical point of view, that sets are in fact a special kind of multisets. In this paper, the definitions of Blizard are adopted and a multiset is a collection of elements, such that each element can occur more than once. The number of occurrences of an element is called the multiplicity of the element. Given a universe U and a multiset A consisting of elements from U, the characteristic function $\omega_A : U \to \mathbb{N}$ of A maps each element to it's multiplicity. For element containment, the following operators are used:

$$\begin{aligned} x \in A \quad &\Leftrightarrow \omega_A(x) > 0 \\ x \notin A \quad &\Leftrightarrow \omega_A(x) = 0 \\ x \in^n A &\Leftrightarrow \omega_A(x) = n \end{aligned} \tag{51}$$

Based on the characteristic function it is possible to define the following operators:

$$\begin{aligned} &\forall u \in U : \omega_{A \cup B}(u) = \max\left(\omega_A(u), \omega_B(u)\right) \\ &\forall u \in U : \omega_{A \cap B}(u) = \min\left(\omega_A(u), \omega_B(u)\right) \\ &\forall u \in U : \omega_{A \uplus B}(u) = \omega_A(u) + \omega_B(u) \end{aligned} \tag{52}$$

The cardinality of a multiset is the total number of elements ($|A| = \sum_{u \in U} \omega_A(u)$) and subsets are defined as $A \subseteq B \Leftrightarrow \forall u \in U : \omega_A(u) \leq \omega_B(u)$. In the case of multisets, the complement operator can not be defined trivially. With regular sets, the complement of a set A is defined as the set difference between the universe and A. The complement is hereby always defined because $A \subset U$ by definition. However, a multiset is not bound to be a subset of the universe U. Several solutions for this problem have been proposed. Chakrabarty[16] assumes that it is always possible to find a largest multiset, which is called the universal multiset. Jena[17] further developed this idea by defining an upper limit n for the multiplicity of elements. The universal multiset is then U_n with $\forall u \in U : \omega_{U_n}(u) = n$. The same problem occurs when defining fuzzy measures for multisets, which is required if the comparison indices are to be defined for multisets. In this paper, the following approach is used. Given a universe U and a multiset U' over U. The set of all sub-multisets of U', including $\emptyset$ and U' itself, is called the multiset space $\mathcal{M}(U')$. If there exists a natural number n such that $\forall u \in U : \omega_{U'}(u) = n$, then $\mathcal{M}(U')$ is called a uniform multiset space, which is denoted by $\mathcal{M}_n(U)$. Using this notation, $\mathcal{M}_\infty(U)$ represents the set of all multisets.

4.2. *Fuzzy multisets*

As regular sets can be generalized to fuzzy sets, so can multisets be generalized to *fuzzy* multisets. Fuzzy multisets are introduced by Yager,[13] who proposed the following characteristic function for a fuzzy multiset. Let $\tilde{\mathcal{M}}_\infty(U)$ denote the set of all fuzzy multisets drawn from universe U, then $A \in \tilde{\mathcal{M}}_\infty(U)$ is characterized by:

$$\psi_A : U \to \mathcal{M}_\infty\left([0,1]\right) : x \mapsto \psi_A(x) \tag{53}$$

With this characteristic function, a multiset over a universe U is defined as a function from the elements of U to regular multisets over the unit interval. Because the characterization by Yager yields difficult notations for operators, other characterizations have been proposed.[16,18,19] One of them is due to Rocacher.[20] Rocacher defines the generalized α-cut of a fuzzy multiset A as a multiset A_α of elements that belong to A with a membership degree of at least α:

$$\forall A \in \tilde{\mathcal{M}}_\infty(U) : \forall \alpha \in \,]\,0,1\,] : \forall u \in U : \omega_{A_\alpha}(u) = \sum_{d \geq \alpha} \omega_{\psi_A(x)}(d) \tag{54}$$

The family of α-cuts $(A_\alpha)_{\alpha \in \,]0,1\,]}$ is a unique characterization of fuzzy multisets. Similarly, an ω-cut of a fuzzy multiset A is defined as a fuzzy set A^ω that contains all elements with multiplicity at least ω:

$$\forall A \in \tilde{\mathcal{M}}_\infty(U) : \forall \omega \in \mathbb{N}_0 : \forall u \in U : \mu_{A^\omega}(u) = \sup\{\alpha | \omega_{A_\alpha} \geq \omega\} \tag{55}$$

The family of ω-cuts $(A^{\omega})_{\omega\in\mathbb{N}_0}$ is a unique characterization of fuzzy multisets. We introduce a new characterization of fuzzy multisets that will be used in the remainder of the paper (Sec. 4.3). A fuzzy multiset A drawn from a universe U is characterized as a regular multiset drawn from $U \times [0,1]$. Hence, $\tilde{\mathcal{M}}_{\infty}(U)$ is an isomorphism of $\mathcal{M}_{\infty}(U \times [0,1])$. This characterization models a fuzzy multiset as a multiset of couples, where each couple consists of an element of U combined with its membership degree.

4.3. *Fuzzy multiset evaluators and multirelations*

With the concept of multiset spaces at hand, it is possible to give a formal definition of scalar fuzzy multiset evaluators:

Definition 3. A function $g : \tilde{\mathcal{M}}(U') \to [0,1]$ is a fuzzy evaluator for the space of fuzzy multisets $\tilde{\mathcal{M}}(U')$ if and only if it satisfies the following properties:

$$\begin{aligned} &\bullet\ g(\emptyset) = 0 \\ &\bullet\ g(U') = 1 \\ &\bullet\ A \subseteq B \Rightarrow g(A) \leq g(B) \end{aligned} \tag{56}$$

A fuzzy measure for multisets is a fuzzy evaluator with the domain limited to regular multisets. Next to evaluators, a second important aspect of the generalized comparison indices is the relation R_s. As relations are sets, the concept of a relation can be generalized to the concept of a *multirelation*, which is, as far as the authors know, not introduced anywhere in literature. A binary relation between two sets A and B is a subset of the cartesian product $A \times B$. Therefor, in order to define a multirelation, the cartesian product of multiset is first defined:

Definition 4. Let A and B be two multisets with universes U and V. The cartesian product $A \times B$ is a multiset with universe $U \times V$. The characteristic function of $A \times B$ is given by:

$$\omega_{A\times B} : U \times V \to \mathbb{N} : (a,b) \mapsto \omega_{A\times B}(a,b) = \omega_A(a)\omega_B(b) \tag{57}$$

Based on Definition 4, the definition of a multirelation is given:

Definition 5. A multirelation between two multisets A and B is a subset of $A \times B$

A multirelation $R \subseteq A \times B$ is called functional if the following two properties are satisfied:

$$\begin{aligned} &\bullet\ \forall (a,b) \in R : \omega_R(a,b) = \omega_A(a) \\ &\bullet\ ((a,b) \in R \wedge (a,b') \in R) \Rightarrow (b = b') \end{aligned} \tag{58}$$

A functional multirelation $R \subseteq A \times B$ is *injective* if:

$$((a,b) \in R \wedge (a',b) \in R) \Rightarrow (a = a') \tag{59}$$

A functional multirelation $R \subseteq A \times B$ is *surjective* if:

$$\forall b \in B : \omega_B(b) \leq \sum_{(a,b)\in R} \omega_A(a) \tag{60}$$

A *bijective* multirelation is a functional multirelation that is both injective and surjective. A one-to-one multirelation is a relation for which each element of one set, is connected to at most one element of the other set:

$$\left(\forall a \in A : \omega_A(a) \geq \sum_{b\in B} \omega_R(a,b)\right) \wedge \left(\forall b \in B : \omega_B(b) \geq \sum_{a\in A} \omega_R(a,b)\right) \tag{61}$$

A complete one-to-one relation is a one-to-one relation such that: $|R| = |A| = |B|$. A complete one-to-one relation that is functional, is a bijection.

4.4. *Comparison indices*

The concepts of (i) a fuzzy evaluator for (fuzzy) multisets and (ii) a multirelation can now be used to define generalized comparison indices for multisets. Only the case of similarity indices is treated here. Inclusion indices and partial matching indices for (fuzzy) multisets can be obtained in a similar manner. Several approaches to define comparison indices for multisets are possible and the preferred approach depends on the application. The first and most obvious approach is to consider different instances of the same element as separate elements. Similarity indices based on Eq. (10) for multisets can be obtained by using generalized definitions of intersection, union (and the derived Δ) en fuzzy evaluators. Extensions of similarity indices that are based on a similarity measure s for elements are obtained by using a multirelation instead of a regular relation. Given two fuzzy multisets A and B in the multiset space $\mathcal{M}(U')$ over U and a similarity measure s over U, it is possible to construct a one-to-one multirelation R_s in $(A \times [0,1]) \times (B \times [0,1])$ such that:

$$C = \sum_{(a,b)\in R_s} T\left(\mu_A(a), \mu_B(b)\right) s(a,b) \tag{62}$$

is maximized. Hereby, T is a t-norm. Note that R_s is not a fuzzy multirelation. Instead, the new characterization of fuzzy multisets is used for elegant notation of the amount of common elements. As R_s is a one-to-one relation, each combination of an element with a membership degree in a fuzzy multiset, is taken into account one time maximally. The amount of common elements is now reflected by C instead of the cardinality of R_s. The similarity of two fuzzy multisets A and B is then calculated as:

$$S_s(A,B) = \frac{C}{|A| + |B| - C} \tag{63}$$

This comparison method is suitable if each element of a multiset can be considered independent, even if some elements are identical. The multiplicity of elements is not explicitly taken into account when calculating the similarity of two multisets. The following example illustrates this:

Example 5. Assume a universe $U = \{a, b, c\}$ on which three multisets are defined: $A = \{a, b, b, b, b\}$, $B = \{a, a, a, b, b\}$ and $C = \{a, b, b, c, c\}$. Let s be a similarity measure over U such that $s(a, b) = s(b, c) = \lambda$. For calculation of $S_s(A, B)$, the optimal one-to-one relation is $R_{s,1} = \{(a, a)_{/1}, (b, a)_{/\lambda}, (b, a)_{/\lambda}, (b, b)_{/1}, (b, b)_{/1}\}$. For calculation of $S_s(A, C)$, the optimal one-to-one relation is $R_{s,2} = \{(a, a)_{/1}, (b, c)_{/\lambda}, (b, c)_{/\lambda}, (b, b)_{/1}, (b, b)_{/1}\}$. Application of Eq. (63) yields that $S_s(A, B) = S_s(A, C) = \frac{3+2\lambda}{7-2\lambda}$.

The second way of constructing comparison indices is to take the multiplicity of elements into account for calculation of multiset similarity. In that case, identical elements are considered as a group and the multiplicity determines the size of the group. Let us first study the case of regular multisets and next the case of fuzzy multisets. With regular multisets, it follows that the multiplicity must be taken into account when calculating the similarity between elements. In order to achieve this, a similarity measure over $U \times \mathbb{N}$ is constructed. Such a similarity measure can be constructed based on a similarity measure s_1 over U and a similarity measure s_2 over $\mathbb{N}$:

$$s\left(\left(a, \omega_A\left(a\right)\right), \left(b, \omega_B\left(b\right)\right)\right) = f\left(s_1\left(a, b\right), s_2\left(\omega_A\left(a\right), \omega_B\left(b\right)\right)\right) \tag{64}$$

with f a monotonic increasing function such that $f(0, 0) = 0$ and $f(1, 1) = 1$. Triangular norms are good choices for f, but use of a uninorm or an averaging function is also possible. Calculation of the similarity of two multisets A and B over U is then implemented by using the comparison method for sets taking into account a similarity measure for elements (Eq. (39)). The comparison of multisets is translated to the comparison of sets $\{(a, \omega_A(a)) | a \in A\}$ and $\{(b, \omega_B(b)) | b \in B\}$.

Example 6. Assume the same setting of Example 5. The multisets can be denoted as $A = \{(a, 1), (b, 4)\}$, $B = \{(a, 3), (b, 2)\}$ and $C = \{(a, 1), (b, 2), (c, 2)\}$. Assume a similarity measure s_2 over $\mathbb{N}$ such that:

$$s_2(n, m) = \frac{1}{1 + |n - m|} \tag{65}$$

and let f be the product t-norm. If it is assumed that $\lambda > 2/3$, then for the calculation of $S_s(A, B)$, the optimal relation is $R_{s,1} = \{((a, 1), (b, 2))_{/\lambda/2}, ((b, 4), (a, 3))_{/\lambda/2}\}$. For the calculation of $S_s(A, C)$, the optimal relation is $R_{s,2} = \{((a, 1), (a, 1))_{/1}, ((b, 4), (b, 2))_{/1/3}\}$. Application of Eq. (39) yields:

$$S_s(A, B) = \frac{|R_{s,1}|}{|A| + |B| - |R_{s,1}|} = \frac{\lambda}{4 - \lambda} \tag{66}$$

$$S_s(A, C) = \frac{|R_{s,2}|}{|A| + |C| - |R_{s,2}|} = \frac{4}{11} \tag{67}$$

which means that $S_s(A,B) \neq S_s(A,C)$ due to the fact that $\lambda \leq 1$. This example thus illustrates how the multiplicity is taken into account to calculate multiset similarity.

For the case of fuzzy multisets, Eq. (64) is generalized by replacing multiplicity of an element a with the sum of membership degrees of a. Formally, with A and B two fuzzy multisets, s_1 a similarity measure over U and s_2 a similarity measure over $\mathbb{R}$:

$$s'\left(\left(a, \sum \mu_A(a)\right), \left(b, \sum \mu_B(b)\right)\right) = f\left(s_1(a,b), s_2\left(\sum \mu_A(a), \sum \mu_B(b)\right)\right). \quad (68)$$

With this similarity measure s', fuzzy multisets are treated as L-fuzzy sets. More specific, the different membership degrees are summed up to one membership degree which is not bound to be an element of $[0,1]$.

5. Conclusion

Comparison of collections is a research topic with many applications in diverse areas of expertise. This paper focuses on the generalization of three types of comparison indices for fuzzy sets: inclusion indices, partial matching indices and similarity indices. The generalization allows to account for similarities between elements. It is shown how fuzzy quantifiers can be used to define comparison indices. Next, the comparison indices are also defined for (fuzzy) multisets. Therefor, fuzzy evaluators for multisets and multirelations are introduced. An alternative characterization of fuzzy multisets is required to allow construction of comparison indices for fuzzy multisets. Two approaches for multiset comparison are considered. In the first approach, each instance of an element is considered as a separate element, which occurs independently of other instances. The second approach considers groups of element instances and the groups are compared, rather than single element instances. In the case of fuzzy multisets, the membership degrees are summed up to one membership degree, possibly not an element of $[0,1]$. Hence, fuzzy multisets are reduced to L-fuzzy sets.

References

1. I. Fellegi and A. Sunter, A theory for record linkage, *American Statistical Association Journal* **64** (1969) 1183–1210.
2. S. Tejada, C. Knoblock and S. Minton, Learning object identification rules for information integration, *Information Systems* **26** (2001) 607–633.
3. Nicolás Marín, J. Medina, O. Pons, D. Sanchez and M. Villa, Complex object comparison in a fuzzy context, *Information and Software Technology* **45** (2003) 431–444.
4. A. Hallez and G. De Tré, *A Hierarchical Approach to Object Comparison*, Lecture Notes in Artificial Intelligence, Canvun, Mexico, 2007, pp. 191–198.
5. D. Dubois and H. Prade, A unifying view of comparison indices in a fuzzy set-theoretic framework, in *Recent Developments in Fuzzy Set and Possibility Theory* (Pergamon Press, 1982) pp. 3–13.
6. P. Jaccard, Nouvelles recherches sur la distribution florale, *Bulletin de la Société de Vaud des Sciences Naturelles* (1908) 44–223.

7. A. Tversky, Features of similarities, *Psychological Review* **84** (1977) 327–352.
8. V. Cross and T. Sudkamp, *Similarity and Compatibility in Fuzzy Set Theory: Assessment and Applications* (Physica-Verlag, 2002).
9. F. Baulieu, A classification of presence/absence based dissimilarity coefficients, *J. Classification* **6** (1989) 233–246.
10. V. Batagelj and M. Bren, Comparing resemblance measures, *J. Classification* **12** (1995) 73–90.
11. M. Rifqi, M. Lesot and M. Detyniecki, Fuzzy order-equivalence for similarity measures, in *Proc. NAFIPS Conference*, New York, USA, 2008, pp. 18–24.
12. L. Zadeh, A computational approach on fuzzy quantifiers in natural languages, *Computing and Mathematics with Applications* **9** (1983) 149–184.
13. R. Yager, On the theory of bags, *Int. J. General Systems* **13** (1986) 23–27.
14. W. Blizard, Generalizations of the Concept of Set: A Formal Theory of Multisets, Ph.D. Thesis, University of Oxford, 1986.
15. W. Blizard, Multiset Theory, *Notre Dame J. Formal Logic* **30** (1989).
16. K. Chakrabarty, R. Biswas and S. Nanda, On Yager's Theory of Bags and Fuzzy Bags, *Computers and Artificial Intelligence* **18** (1999), 7–17
17. S. Jena, S. Ghosh and B. Tripathy, On the theory of bags and lists, *Information Sciences: an International Journal* **132** (2001) 241–254
18. A. Rebai, Canonical fuzzy bags and bag fuzzy measures as a basis for madm with mixed non cardinal data, *European J. Operational Research* **78** (1994) 34–48
19. A. Rebai and J. Martel, A fuzzy bag approach to choosing the best multiattributed potential actions in multiple judgement and non cardinal data context, *Fuzzy Sets and Systems* **87** (1997) 159–166
20. D. Rocacher, On fuzzy bags and their application to flexible querying, *Fuzzy Sets and Systems* **140** (2003) 93–110