

# Uncertainty-aware similarity measures - properties and construction method

Patryk Żywica and Anna Stachowiak

Department of Imprecise Information Processing Methods,  
Faculty of Mathematics and Computer Science,  
Adam Mickiewicz University in Poznań,  
Umultowska 87, 61-614 Poznań, Poland,  
bikol@amu.edu.pl, aniap@amu.edu.pl

## Abstract

A concept of *uncertainty-aware* similarity measure is being defined and discussed. The aim of the paper is to support the opinion that both definition and construction of such measure should take into account the epistemic nature of compared incomplete (uncertain) information. This approach, however, generates new challenges resulted from the computational complexity of the problem. We define a set of properties to be satisfied by *uncertainty-aware* similarity measure and we propose a new technique of constructing such measures for Interval-Valued Fuzzy Sets.

**Keywords:** Similarity, Uncertainty, Interval-Valued Fuzzy Sets, IVFS.

## 1 Introduction

The need for modeling imprecise and incomplete information gave rise to a theory of fuzzy sets and its many extensions. In this paper we give a special attention to Interval-Valued Fuzzy Sets (IVFS) since they provide a way not only to model vagueness of membership values (information imprecision) but also hesitation about those values (information incompleteness or uncertainty).

An IVFS can be defined as a set of possible fuzzy sets, one of which is the "true" or "real" one, presently not known due to the lack of knowledge. Thus, IVFS is a way to describe or represent some information that is uncertain (the uncertainty is not of the probability type, but arises from the lack of knowledge; it can be reduced when knowledge increases). Such interpretation of an interval is of epistemic nature - contrary to ontic one, when the interval is understood as a complex, but certain, information [7]. We would thus refer to an IVFS as to the set of its possible states. We

clearly distinguish a description of an (unknown) object from the object itself (represented by one of the possible states). The notion of *uncertainty-aware* similarity measure that is proposed in this paper takes this distinction into account. Let us assume for example that we want to compare two identical IVFSs -  $[0.1, 0.8]$  and  $[0.1, 0.8]$ ; obviously we notice the total similarity of their description, however, it does not imply total similarity of the objects that are being described.

The paper is entirely devoted to considerations about *uncertainty-aware* similarity measures that take into account an epistemic nature of data and their construction methods for IVFSs. We propose a set of properties to be satisfied by such measure, and we give a method to construct it. The motivation for our work is clear - similarity measure plays a fundamental role in many fields and applications such as approximate reasoning, decision-making systems, recommender systems, pattern recognition and others. On the other hand we believe that the problem of information incompleteness still requires closer look and insightful investigation.

## 2 Definitions

Let  $U = \{u_1, u_2, \dots, u_n\}$  be a crisp universal set. A mapping  $A : U \rightarrow [0, 1]$  is called a fuzzy set (FS) in  $U$ . For each  $1 \leq i \leq n$ , the value  $A(u_i)$  ( $a_i$  for short) represents the membership grade of  $u_i$  in  $A$ . Any crisp set  $X \subseteq U$  can be represented as a fuzzy set by its characteristic function  $\mathbb{1}_X$ . Let  $\mathcal{F}(U)$  be the family of all fuzzy sets in  $U$ .

A binary operation  $t : [0, 1] \times [0, 1] \rightarrow [0, 1]$  is called a triangular norm (t-norm, for short) if it is commutative, associative, non-decreasing in each argument, and has 1 as neutral element. The most important t-norms are minimum  $t_{\min}(x, y) = \min(x, y)$ , product  $t_{\text{prod}}(x, y) = xy$ , and Łukasiewicz  $t_{\text{Luk}}(x, y) = \max(0, x + y - 1)$ . A thorough investigation on t-norms is done in the classical monograph of Klement

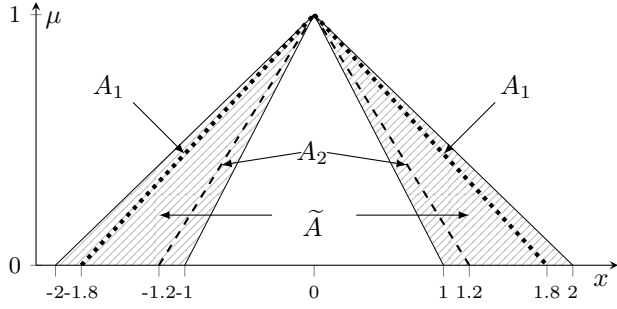


Figure 1: Visualization of the Interval-Valued Fuzzy Set  $\tilde{A}$  and its Fuzzy Membership Function Family  $\tilde{A}$  (shaded area). Dashed lines represent two embedded fuzzy sets (possible states)  $A_1 \in \tilde{A}$  and  $A_2 \in \tilde{A}$ .

et. al.[13].

**Definition 1.** A similarity measure of fuzzy sets is defined as a function on  $\mathbb{E} \subset \mathcal{F}(U) \times \mathcal{F}(U)$  [6, 30]

$$s : \mathbb{E} \rightarrow \mathbb{R}, \quad (1)$$

where  $\mathbb{E}$  needs to satisfy:

1.  $(A, B) \in \mathbb{E}$  if and only if  $(B, A) \in \mathbb{E}$ ,
2.  $(A, B) \in \mathbb{E}$  if and only if  $(A, \mathbb{1}_U) \in \mathbb{E}$ .

It is common to assume that the higher measure values indicate higher similarity of arguments.

In the case where  $\mathbb{E} = \mathcal{F}(U) \times \mathcal{F}(U)$ , all fuzzy sets are comparable by a given similarity measure. Such a situation is not always possible, because some of the standard similarity measures are not defined for certain pairs of fuzzy sets.

Any closed subset  $\tilde{A}$  of  $\mathcal{F}(U)$  will be called Fuzzy Membership Function Family (FMFF) (see [35]). Set  $\tilde{A}$  represents all the possible states that can hide behind uncertain information. Example FMFF is given on Figure 1. We denote by  $\mathcal{FMFF}(U)$  the set of all FMFFs. This approach is largely inspired by the Mendel representation theorem and his Wavy-Slice representation [16, 17].

The cardinality of fuzzy sets has been extensively discussed in the literature (see [27]). In this paper we will focus on scalar cardinalities of fuzzy sets which can be characterised by the formula

$$\sigma_f(A) = \sum_{1 \leq i \leq n} f(A(x_i)), \quad (2)$$

where  $f : [0, 1] \rightarrow [0, 1]$  is a weighting function such that  $f(0) = 0$ ,  $f(1) = 1$  and  $f(a) \leq f(b)$  whenever  $a \leq$

$b$ . This approach formalises and reflects real human counting process under information imprecision [28]. The most common weighting function is the identity function  $f_{id}(x) = x$ .

### 3 Similarity

Defining the similarity of epistemic data is a complex problem. For example, it is necessary to answer the question of how to determine the degree of similarity, so that it reflects the similarity of information described in an incomplete way. As it was noted in the Introduction, even the total similarity of incomplete descriptions does not guarantee the similarity of the described phenomena or objects. For this reason, it is necessary to model the similarity by means of a range or subset.

Thanks to collecting and systematising properties of many similarity measures [2–5, 8, 11, 14, 15, 18, 20–24, 29, 32–34], it is possible to propose the concept of *uncertainty-aware* similarity measure. In the further part of the paper, examples of such measures will be presented along with their basic properties.

Since the similarity of uncertain objects is a more complex problem in comparison to classical similarity measure, the set of properties to be hold is wider.

**Definition 2.** A function  $\tilde{s} : \tilde{\mathbb{E}} \rightarrow \mathcal{P}([0, 1])$  on  $\tilde{\mathbb{E}} \subset \mathcal{FMFF}(U) \times \mathcal{FMFF}(U)$  such that:

1.  $(\tilde{A}, \tilde{B}) \in \tilde{\mathbb{E}}$  if and only if  $(\tilde{B}, \tilde{A}) \in \tilde{\mathbb{E}}$ ,
  2.  $(\tilde{A}, \tilde{B}) \in \tilde{\mathbb{E}}$  if and only if  $(\tilde{A}, \{\mathbb{1}_U\}) \in \tilde{\mathbb{E}}$ ,
  3.  $(\tilde{A}, \tilde{B}) \in \tilde{\mathbb{E}}$  if and only if for any fuzzy sets  $A \in \tilde{A}$ ,  $B \in \tilde{B}$ :
- $$(\{A\}, \{B\}) \in \tilde{\mathbb{E}}, \quad (3)$$

is a *uncertainty-aware* similarity measure if it satisfies following conditions:

(P1) For all  $(\tilde{A}, \tilde{B}) \in \tilde{\mathbb{E}}$ ,

$$\tilde{s}(\tilde{A}, \tilde{B}) = \tilde{s}(\tilde{B}, \tilde{A}). \quad (4)$$

(P2) For  $\tilde{A}, \tilde{B} = \mathcal{F}(U)$ ,

$$\tilde{s}(\tilde{A}, \tilde{B}) = [0, 1] \quad (5)$$

(P3) For all  $(\tilde{A}, \tilde{B}) \in \tilde{\mathbb{E}}$ ,  $(\tilde{A}, \tilde{C}) \in \tilde{\mathbb{E}}$  such that  $\mathbb{1}_X \in \tilde{A}$ ,  $\mathbb{1}_X \in \tilde{B}$  and  $\mathbb{1}_{X^c} \in \tilde{C}$  for some  $X \subset U$ ,

$$1 \in \tilde{s}(\tilde{A}, \tilde{B}), \quad (6)$$

$$0 \in \tilde{s}(\tilde{A}, \tilde{C}). \quad (7)$$

(P4) For all fuzzy sets  $A, B \in \mathcal{F}(U)$  such that  $(\{A\}, \{B\}) \in \tilde{\mathbb{E}}$ ,

$$\tilde{s}(\{A\}, \{B\}) = \{a\}, \text{ for some } a \in [0, 1]. \quad (8)$$

(P5) For all  $(\tilde{A}, \tilde{B}) \in \tilde{\mathbb{E}}$  and for any  $A, B \in \mathcal{F}(U)$  such that  $A \in \tilde{A}, B \in \tilde{B}$ ,

$$\tilde{s}(\{A\}, \{B\}) \subset \tilde{s}(\tilde{A}, \tilde{B}). \quad (9)$$

(P6) For any  $(\tilde{A}, \tilde{C}) \in \tilde{\mathbb{E}}, (\tilde{B}, \tilde{D}) \in \tilde{\mathbb{E}}$  such that  $\tilde{A} \subset \tilde{B}$  and  $\tilde{C} \subset \tilde{D}$ ,

$$\tilde{s}(\tilde{A}, \tilde{C}) \subset \tilde{s}(\tilde{B}, \tilde{D}). \quad (10)$$

(P7) For all  $(\tilde{A}, \tilde{B}) \in \tilde{\mathbb{E}}$  such that  $0 \in \tilde{s}(\tilde{A}, \tilde{B})$  there exist  $A \in \tilde{A}$  and  $B \in \tilde{B}$  such that  $\sigma(A \cap B) = 0$ .

To support the choice of the properties (P1)-(P7) let us now discuss them briefly.

The first property, symmetry, is a common and widely accepted condition for every similarity measure, and so it is in the presence of uncertainty. Next properties should be considered taking into account the specificity of epistemic information. Thus, (P2) requires that no information implies no conclusions - when comparing totally unknown object, the similarity should also remain unknown. On the other hand, if the information is complete (FMFF reduces to a single FS) then their similarity should also be completely known (without uncertainty) - that is the meaning of the property (P4).

By (P3) we make two observations; two FMFFs could be similar to a degree 1 only if they share at least one common state (6). On the other hand, if two FMFFs are to some extent inconsistent, then a value 0 should be a possible value of their similarity (7). In general, when a degree of uncertainty of two FMFFs decreases so does similarity measure - see (P6). Consequently, for any pair of possible states, their similarity measure belongs to the similarity of FMFFs (P5).

Finally, (P7) indicates that if a value 0 is one of the possible values of similarity of two FMFFs then there exists two states that are disjoint.

#### 4 Uncertainty-aware similarity measures for IVFS

Interval-valued fuzzy set (IVFS) theory, which is a special case of type-2 fuzzy set theory, was introduced by Zadeh [31]. Let  $\mathcal{I}([0, 1])$  be the set of all closed subintervals of  $[0, 1]$ . A mapping  $\hat{A} : U \rightarrow \mathcal{I}([0, 1])$  is called an interval-valued fuzzy set. For each  $1 \leq i \leq n$ ,

the value  $\hat{A}(u_i) = [\underline{A}(u_i), \overline{A}(u_i)] \in \mathcal{I}([0, 1])$  represents the membership of an element  $u_i$  in  $\hat{A}$ . Usually  $\underline{A}$  and  $\overline{A}$  are called the lower and upper membership functions of  $\hat{A}$  respectively. In epistemic approach, interval  $\hat{A}(u_i)$  is understood to contain the true membership degree of  $u_i$  in some incompletely known fuzzy set  $A$  represented by  $\hat{A}$ . We denote the set of all interval-valued fuzzy sets in  $U$  by  $\mathcal{IV}(U)$ .

Most known extensions of fuzzy sets can be fully accurately represented using FMFF. Interval-Valued Fuzzy Set  $\hat{A}$  can be also viewed as the following FMFF (see Figure 1):

$$\tilde{A} = \{A \in \mathcal{F}(U) : \forall_{x \in U} \underline{A}(x) \leq A(x) \leq \overline{A}(x)\}. \quad (11)$$

Referring to the Mendel's Wavy-Slice representation theorem [16, 17], in case of Interval-Valued Fuzzy Sets, FMFF is equivalent to FOU (Footprint of Uncertainty)

$$\tilde{A} = \text{FOU}(\hat{A}). \quad (12)$$

Since there is a one-to-one correspondence between  $\hat{A}$  and  $\tilde{A}$ , those two representations will be used interchangeably.

In the previous sections, we presented the properties of a *uncertainty aware* similarity measure for the general epistemic data represented by FMFF. In the following we will propose the construction method of such measures for IVFS. Note that each function  $f : X \rightarrow Y$  can be calculated for set-valued data  $A \subset X$  in following way

$$f(A) = \{f(a) : a \in A\} \subset Y. \quad (13)$$

This approach can be used to obtain new similarity measures for uncertain data.

**Definition 3.** Let  $s : E \rightarrow [0, 1]$  be a similarity measure of fuzzy sets. Function  $\hat{s} : \tilde{\mathbb{E}} \rightarrow \mathcal{P}([0, 1])$  can be defined in following way:

$$\hat{s}(\hat{A}, \hat{B}) = \{s(A, B) : A \in \tilde{A}, B \in \tilde{B}\}, \quad (14)$$

where

$$\tilde{\mathbb{E}} = \{(\hat{A}, \hat{B}) \in \mathcal{IV}(U) \times \mathcal{IV}(U) : \tilde{A} \times \tilde{B} \subset \mathbb{E}\}. \quad (15)$$

Sometimes it may be useful to represent fuzzy set  $A$  on finite universe  $U$  as a vector  $\mathbf{x}_A = (A(u_1), \dots, A(u_{|U|})) \in [0, 1]^{|U|}$ .

**Definition 4.** Fuzzy set similarity measure  $s : \mathbb{E} \rightarrow [0, 1]$  is called continuous if function  $f : X \rightarrow [0, 1]$  defined on

$$X = \{(\mathbf{x}_A, \mathbf{x}_B) \in [0, 1]^{2|U|} : (A, B) \in \mathbb{E}\}. \quad (16)$$

as

$$f(\mathbf{x}_A, \mathbf{x}_B) = s(A, B), \quad (17)$$

is continuous in the whole domain.

**Lemma 1.** For each continuous fuzzy set similarity measure  $s : \mathbb{E} \rightarrow [0, 1]$ , function  $\hat{s}$  from Definition 3 can be simplified to

$$\hat{s}(\hat{A}, \hat{B}) = \left[ \inf_{\substack{A \in \hat{A} \\ B \in \hat{B}}} s(A, B), \sup_{\substack{A \in \hat{A} \\ B \in \hat{B}}} s(A, B) \right]. \quad (18)$$

**Theorem 2.** For each continuous fuzzy set similarity measure  $s : \mathbb{E} \rightarrow [0, 1]$  that satisfies

1. for each  $(A, B) \in \mathbb{E}$ , we have  $s(A, B) = s(B, A)$ ,
2. for each  $(A, B) \in \mathbb{E}$ , if  $s(A, B) = 0$  then  $\sigma(A \cap B) = 0$ ,
3. for each  $X \subset U$  such that  $(\mathbf{1}_X, \mathbf{1}_{X^c}) \in \mathbb{E}$  we have  $(\mathbf{1}_X, \mathbf{1}_{X^c}) = 0$  and  $(\mathbf{1}_X, \mathbf{1}_X) = 1$ ,

function  $\hat{s}$  from Definition 3 is uncertainty-aware similarity measure of IVFS from Definition 2.

The proof was given in [34].

## 5 Extensions of popular similarity measures

The following section presents the extensions of known similarity measures to their *uncertainty-aware* versions obtained using the Definition 3. As will be shown, the calculation of some generated similarity measures is computationally difficult, while other measures can be calculated using simple formulas. A particularly interesting case is the Jaccard index.

### 5.1 Distance based similarity measures

The similarity measure based on the metric of the distance meets the assumption of the Theorem 2. Thanks to this, appropriate extensions are *uncertainty-aware* similarity measures.

#### 5.1.1 Minkowski distance

The extensions of similarity measures based on generalized Minkowski's metric are relatively simple to calculate. By using (18) we get the following measure of similarity:

$$\hat{s}_{d_r}(\hat{A}, \hat{B}) = \left[ \inf_{\substack{A \in \hat{A} \\ B \in \hat{B}}} 1 - \frac{d_r(A, B)}{|U|}, \sup_{\substack{A \in \hat{A} \\ B \in \hat{B}}} 1 - \frac{d_r(A, B)}{|U|} \right], \quad (19)$$

which is defined for all pairs of interval-valued fuzzy sets. For simplicity, let us denote this interval by  $[a, b]$ .

Then we can transform the lower limit value in the following way:

$$\begin{aligned} a &= \inf_{\substack{A \in \hat{A} \\ B \in \hat{B}}} 1 - \frac{d_r(A, B)}{|U|} = 1 - \sup_{\substack{A \in \hat{A} \\ B \in \hat{B}}} \frac{d_r(A, B)}{|U|} \\ &= 1 - \frac{1}{|U|} \sup_{\substack{u \in U \\ a_u \in \hat{A}(u) \\ b_u \in \hat{B}(u)}} \left( \sum_{u \in U} |a_u - b_u|^r \right)^{\frac{1}{r}} \\ &= 1 - \frac{1}{|U|} \left( \sum_{u \in U} \max \{ |\underline{A}(u) - \underline{B}(u)|, |\overline{A}(u) - \overline{B}(u)| \}^r \right)^{\frac{1}{r}}. \end{aligned} \quad (20)$$

Analogous transformations lead to a direct formula for the upper limit  $b$ . This formula is a simple sum and no numerical optimisation is needed to calculate it. Another important observation is the fact that this formula takes into account only values of the lower and upper membership functions of IVFSs  $\hat{A}$  and  $\hat{B}$ . Hence, these measures are computationally efficient despite the fact that their construction uses a computationally inefficient method that takes into account the infinite FMFF.

In a very similar way we can get extensions of similarity measures based on  $d_\infty$  distance. The formulas obtained are analogous and have the same computational complexity.

#### 5.1.2 Other distances

The similarity measure can also be defined as the angle between two vectors:

$$s_{\cos \theta}(A, B) = \frac{\sum_{u \in U} \mu_A(u) \mu_B(u)}{\sqrt{\sum_{u \in U} \mu_A(u)^2} \sqrt{\sum_{u \in U} \mu_B(u)^2}}. \quad (21)$$

In the case when any of the fuzzy sets is empty, the measure is not specified, so

$$\mathbb{E} = \{(A, B) \in \mathcal{IV}(U) \times \mathcal{IV}(U) : A \neq \mathbf{1}_\emptyset \text{ \& } B \neq \mathbf{1}_\emptyset\}. \quad (22)$$

Measure  $s_{\cos \theta}(A, B)$  takes the largest value when the angle between the vectors representing the fuzzy sets is  $90^\circ$ . This similarity measure is used when it is particularly important to compare the shape of the membership function of fuzzy sets. The cosine of the angle is not the only angular measure. Many modifications of (21) can be found in the literature [12].

The above definition can not easily be simplified as it was in the case of Minkowski's distance. The problem of constructing an algorithm that enables the efficient calculation of such a similarity measure is still open.

## 5.2 Logic based similarity measures

Logic-based measures [9, 10] use the interpretation of the membership function of a fuzzy set as the degree of truth of the proposition represented by this fuzzy set. The basic method assumes the use of an implication operator, which allows constructing both the measures of inclusion and similarity. In classic logic, the implication operator can be defined in several equivalent ways. The generalization to the case of fuzzy logic, where infinitely many degrees of truth are admitted, resulted in the creation of many not equivalent definitions of the concept. The most frequently used implication operators are S-implications and R-implications [1, 25].

The simplest co-implication operator is defined as:

$$\Psi(a, b) = \min(a \Rightarrow b, b \Rightarrow a). \quad (23)$$

The similarity measure is then defined as the minimum, average or maximum value obtained for all elements of the universe. The most interesting is the case of the average, where the similarity measure is defined as

$$s_{\Psi}(A, B) = \frac{1}{|U|} \sum_{u \in U} \Psi(\mu_A(u), \mu_B(u)). \quad (24)$$

For simplicity, we denote by  $[a, b]$  the similarity value returned by measure extended according to the Definition 3. Then the value of the lower bound can be transformed as follows:

$$\begin{aligned} a &= \inf_{\substack{A \in \tilde{\mathcal{A}} \\ B \in \tilde{\mathcal{B}}}} \frac{1}{|U|} \sum_{u \in U} \Psi(\mu_A(u), \mu_B(u)) = \\ &= \frac{1}{|U|} \sum_{u \in U} \inf_{\substack{\mu_{\hat{A}}(u) \leq x \leq \bar{\mu}_{\hat{A}}(u) \\ \mu_{\hat{B}}(u) \leq y \leq \bar{\mu}_{\hat{B}}(u)}} \Psi(x, y). \end{aligned} \quad (25)$$

Similarly, we can convert the upper limit. This simple transformation allows for a significant simplification of the problem of calculating the infimum and supremum. Instead of optimizing the value of the whole sum, it is enough to examine how the co-implication operator behaves. For example, for the Łukasiewicz's implication operator, we get the following equality:

$$\begin{aligned} &\inf_{\substack{\mu_{\hat{A}}(u) \leq x \leq \bar{\mu}_{\hat{A}}(u) \\ \mu_{\hat{B}}(u) \leq y \leq \bar{\mu}_{\hat{B}}(u)}} \Psi_{\text{Luk}}(x, y) \\ &= \inf_{\substack{\mu_{\hat{A}}(u) \leq x \leq \bar{\mu}_{\hat{A}}(u) \\ \mu_{\hat{B}}(u) \leq y \leq \bar{\mu}_{\hat{B}}(u)}} \min\{1, 1 - x + y, 1 - y + x\} \\ &= \min\{1, 1 - \bar{\mu}_{\hat{A}}(u) + \underline{\mu}_{\hat{B}}(u), 1 - \bar{\mu}_{\hat{B}}(u) + \underline{\mu}_{\hat{A}}(u)\}. \end{aligned}$$

Thanks to this property, the calculation of the extended similarity measure  $\hat{s}_{\Psi}$  is possible directly without the need for numerical optimization techniques.

Unfortunately, the problem of calculating the measure  $\hat{s}_{\Psi}$  is not so simple in the general case. For the two basic families of implication operators: S-implications and R-implications, the simple way to directly calculate the similarity measure is not known. It is an interesting and still open problem for further research, but it is not in the scope of this paper.

## 5.3 Set theory based similarity measures

The Jaccard index is the most commonly used similarity measure. It formalizes the observation that, for two sets: the more common and less different elements they have, the more similar they are. As a reminder, the Jaccard index for fuzzy sets is defined as

$$s_J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \text{ where } |A \cup B| \neq 0. \quad (26)$$

Because  $A \cap B \subset A \cup B$ , Jaccard index can be viewed as the ratio of the number of common elements of  $A$  and  $B$  to the number of all elements in  $A$  or  $B$ . Another look at Jaccard's index comes from the observation that

$$\begin{aligned} s_J(A, B) &= \frac{|A \cap B|}{|A \cup B|} = \frac{|(A \cup B) \cap (A \cap B)|}{|A \cup B|} \\ &= \sigma(A \cap B | A \cup B). \end{aligned} \quad (27)$$

According to the interpretation of relative cardinality as a degree of inclusion, the degree of similarity between two fuzzy sets is defined as the degree to which the fuzzy set  $A \cup B$  is contained in  $A \cap B$ . Of course, the opposite inclusion always holds. This approach is also interesting because it refers to the concept of inclusion of fuzzy sets (in this case defined as relative cardinality).

Both proposed interpretations can be generalized using t-norm  $T$  and t-conorm  $S$ . In addition, the cardinality of the fuzzy set can also be defined using any weighting function  $f$ . In this way, we get the following two definitions of the generalized Jaccard index:

$$s'_{T,S,f}(A, B) = \frac{\sigma_f(A \cap_T B)}{\sigma_f(A \cup_S B)}, \quad (28)$$

$$\begin{aligned} s''_{T,S,f}(A, B) &= \sigma_{T,f}(A \cap_T B | A \cup_S B) \\ &= \frac{\sigma_f((A \cup_S B) \cap_T (A \cap_T B))}{\sigma_f(A \cup_S B)}. \end{aligned} \quad (29)$$

It should be noted that both generalizations are not specified in the case where  $\sigma_f(A \cup_S B) = 0$ . Unfortunately, the unambiguous definition of the similarity value in this case is not possible. Thus, the domain of similarity measures can be defined as follows:

$$\mathbb{E}_{S,f} = \{(A, B) \in \mathcal{F}(U) \times \mathcal{F}(U) : \sigma_f(A \cup_S B) \neq 0\}. \quad (30)$$



For any t-conorm  $S$  and weighting function  $f$ ,  $\mathbb{E}_{S,f}$  satisfies both conditions required for the domain of similarity measure given in the Definition 1. In addition, many specific pairs of fuzzy sets belong to this family. For example, for any set of  $X \subset U$  both  $(\mathbb{1}_X, \mathbb{1}_X) \in \mathbb{E}_{S,f}$ , as well as  $(\mathbb{1}_X, \mathbb{1}_{X^c}) \in \mathbb{E}_{S,f}$ , which results directly from the properties of t-operations and operations on fuzzy sets. One special case seems particularly important when

$$\mathbb{E}_{S,f} = \mathcal{IV}(U) \times \mathcal{IV}(U) \setminus \{(\emptyset, \emptyset)\} = \mathbb{E}_\emptyset, \quad (31)$$

which occurs, for example, for  $\mathbb{E}_{S_{\max}, f_{id}}$ .

Both similarity measures are equivalent for both classic and fuzzy sets, for which the intersection and sum of sets is defined using  $T_{\min}$  and  $S_{\max}$  with the identity weighting function. For the other t-operations this equality may not hold. Therefore, in the further part of this section, we will consider them separately. As it turns out, they have very similar but not identical properties.

**Theorem 3.** *Fuzzy set similarity measure  $s'_{T,S,f}$  meets the assumptions of the Theorem 2. In addition, if the t-norm  $T$  does not have zero divisors and  $\forall_{x \in (0,1]} f(x) > 0$ ,  $s''_{T,S,f}$  also has this property.*

**Remark 1.** *Classic fuzzy Jaccard index*

$$s'_{T_{\min}, S_{\max}, f_{id}}(A, B) = \frac{|A \cap_{\min} B|}{|A \cup_{\max} B|} \quad (32)$$

satisfies the assumptions of Theorem 2.

**Theorem 4.** *Fuzzy set similarity measures  $s'_{T,S,f}$  and  $s''_{T,S,f}$  are continuous if functions  $T$ ,  $S$ ,  $f$  are continuous in their entire domains.*

As with the angular distance, it is not possible to easily simplify formulas for lower and upper limits. Measure  $s''_{T,S,f}$  is defined by the generalized relative cardinality of fuzzy sets, making it possible to simplify the problem of its efficient calculation. For this purpose, we will use the concept of generalized relative cardinality of interval-valued fuzzy sets [36].

**Definition 5.** Function  $\hat{\sigma}_{f,T} : \mathbb{E}_f \rightarrow \mathcal{I}([0, 1])$  induced by a continuous t-norm  $T$  and a continuous weighting function  $f$  where

$$\hat{\sigma}_{f,T}(\hat{A}|\hat{B}) = \left[ \inf_{\substack{A \in \tilde{\hat{A}} \\ B \in \tilde{\hat{B}}}} \sigma_{f,T}(A|B), \sup_{\substack{A \in \tilde{\hat{A}} \\ B \in \tilde{\hat{B}}}} \sigma_{f,T}(A|B) \right]$$

and

$$\mathbb{E}_f = \{(\hat{A}, \hat{B}) \in \mathcal{IV}(U) \times \mathcal{IV}(U) : \forall_{B \in \tilde{\hat{B}}} \sigma_f(B) > 0\}, \quad (33)$$

is called *generalized interval-valued relative cardinality*.

One can see the following relationship between the extended Jaccard's index  $\hat{s}''_{T,S,f}$  and the generalized interval-valued relative cardinality  $\hat{\sigma}_{f,T}$ :

$$\begin{aligned} \hat{s}''_{T,S,f}(\hat{A}, \hat{B}) &= \\ &= \left[ \inf_{\substack{A \in \tilde{\hat{A}} \\ B \in \tilde{\hat{B}} \\ A' = A \cap_T B \\ B' = A \cup_S B}} \sigma_{f,T}(A'|B'), \sup_{\substack{A \in \tilde{\hat{A}} \\ B \in \tilde{\hat{B}} \\ A' = A \cap_T B \\ B' = A \cup_S B}} \sigma_{f,T}(A'|B') \right] \\ &= \hat{\sigma}_{f,T}(\hat{A} \cap_T \hat{B} | \hat{A} \cup_S \hat{B}). \end{aligned} \quad (34)$$

Thanks to the above equation, the value of the extended Jaccard index can be calculated by means of generalized interval-valued relative cardinality. The effective solution of this computational problem was given in [36]. Unfortunately, the problem of effective calculation of the  $s'_{T,S,f}$  in the general case still remains an open problem. However, it should be mentioned that for the extension of the classical Jaccard index, effective algorithms enabling its calculation in  $O(n \log n)$  were given [19, 26].

## 6 Conclusions

There are two main contributions made in this paper. The first one was collecting and systematizing a set of properties for *uncertainty-aware* similarity measure. Our approach gives a full picture of the similarity of incompletely known information and allows reasoning about the amount of uncertainty of compared objects, thus informs about the quality of this comparison. The second result, presented in Section 4 and 5, is the observation, that *uncertainty-aware* similarity measure may be constructed from fuzzy similarity measures under certain conditions. In this way we have opened a new path for constructing new similarity measures. We showed that this path is very promising since we have already managed to obtain some interesting results by restricting our considerations to IVFS. However, it must be mentioned that processing general epistemic data requires much more research to be computationally feasible.

## Acknowledgement

This work was partially supported by the Polish National Science Centre grant number 2016/21/N/ST6/00316.

## References

- [1] M. Baczyński, B. Jayaram, Fuzzy Implications, Springer, Heidelberg, 2008.
- [2] H. Bustince, Indicator of inclusion grade for interval-valued fuzzy sets. application to approximate reasoning based on interval-valued fuzzy

- sets, *International Journal of Approximate Reasoning* 23 (3) (2000) 137–209.
- [3] H. Bustince, P. Burillo, Vague sets are intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 79 (3) (1996) 403–405.
  - [4] S.-M. Chen, Measures of similarity between vague sets, *Fuzzy Sets and Systems* 74 (2) (1995) 217–223.
  - [5] I. Couso, L. Garrido, L. Sánchez, Similarity and dissimilarity measures between fuzzy sets: A formal relational study, *Information Sciences* 229 (2013) 122–141.
  - [6] V. V. Cross, T. A. Sudkamp, *Similarity and Compatibility in Fuzzy Set Theory. Assessment and Applications*, Physica-Verlag, Heidelberg, 2002.
  - [7] D. Dubois, H. Prade, Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets, *Fuzzy Sets and Systems* 192 (2012) 3–24.
  - [8] W.-L. Gau, D. J. Buehrer, Vague sets, *IEEE Transactions on Systems, Man and Cybernetics* 23 (2) (1993) 610–614.
  - [9] K. Hirota, W. Pedrycz, Handling fuzziness and randomness in process of matching fuzzy data, in: *Proceedings of the Third IFSA Congress, Seattle, USA, 1989*, pp. 97–100.
  - [10] K. Hirota, W. Pedrycz, Matching fuzzy quantities, *IEEE Transactions on Systems, Man and Cybernetics* 21 (6) (1991) 1580–1586.
  - [11] I. Jenhani, S. Benferhat, Z. Elouedi, Possibilistic similarity measures, in: B. Bouchon-Meunier, L. Magdalena, et al. (Eds.), *Foundations of Reasoning under Uncertainty*, Springer-Verlag, Heidelberg, 2010, pp. 99–123.
  - [12] T. Kailath, The divergence and Bhattacharyya distance measures in signal selection, *IEEE Transactions on Communication Technology* 15 (1) (1967) 52–60.
  - [13] E. P. Klement, R. Mesiar, E. Pap, *Triangular norms*, Vol. 8 of *Trends in Logic*, Springer, 2000.
  - [14] Z. Liang, P. Shi, Similarity measures on intuitionistic fuzzy sets, *Pattern Recognition Letters* 24 (15) (2003) 2687–2693.
  - [15] X. Luo, C. Zhang, An axiom foundation for uncertain reasonings in rule-based expert systems: NT-algebra, *Knowledge and Information Systems* 1 (4) (1999) 415–433.
  - [16] J. M. Mendel, Tutorial on the uses of the interval type-2 fuzzy set's Wavy Slice Representation Theorem, in: *Proceedings of Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS)*, New York, USA, 2008, pp. 1–6.
  - [17] J. M. Mendel, R. I. John, F. Liu, Interval type-2 fuzzy logic systems made simple, *IEEE Transactions on Fuzzy Systems* 14 (6) (2006) 808–821.
  - [18] H. B. Mitchell, On the Dengfeng–Chuntian similarity measure and its application to pattern recognition, *Pattern Recognition Letters* 24 (16) (2003) 3101–3104.
  - [19] H. T. Nguyen, V. Kreinovich, Computing degrees of subthood and similarity for interval-valued fuzzy sets: fast algorithms, Tech. Rep. 94, Department of Computer Science, UTEP (2008).
  - [20] A. Stachowiak, P. Żywica, K. Dyczkowski, A. Wójtowicz, An Interval-Valued Fuzzy Classifier Based on an Uncertainty-Aware Similarity Measure, in: P. Angelov, K. T. Atanassov, et al. (Eds.), *Intelligent Systems' 2014. Volume 1: Mathematical Foundations, Theory, Analyses*, Springer, Switzerland, 2015, pp. 741–751.
  - [21] E. Szmidt, *Distances and Similarities in Intuitionistic Fuzzy Sets*, Springer, Switzerland, 2014.
  - [22] E. Szmidt, J. Kacprzyk, Distances between intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 114 (3) (2000) 505–518.
  - [23] E. Szmidt, J. Kacprzyk, Entropy for intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 118 (3) (2001) 467–477.
  - [24] E. Szmidt, J. Kacprzyk, A measure of similarity for intuitionistic fuzzy sets, in: *Proceedings of EUSFLAT Conference, Zittau, Niemcy, 2003*, pp. 206–209.
  - [25] E. Trillas, L. Valverde, On mode and implication in approximate reasoning, in: M. Gupta, A. Kandel (Eds.), *Approximate reasoning in expert systems*, North-Holland, Amsterdam, 1985, pp. 157–166.
  - [26] D. Wu, J. M. Mendel, Efficient algorithms for computing a class of subthood and similarity measures for interval type-2 fuzzy sets, in: *Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ)*, Barcelona, Hiszpania, 2010, pp. 1–7.
  - [27] M. Wygralak, *Cardinalities of fuzzy sets*, Springer, Heidelberg, 2003.
  - [28] M. Wygralak, *Intelligent Counting Under Information Imprecision*, Vol. 292 of *Studies in Fuzziness and Soft Computing*, Springer, 2013.
  - [29] L. Xuecheng, Entropy, distance measure and similarity measure of fuzzy sets and their relations, *Fuzzy Sets and Systems* 52 (3) (1992) 305–318.
  - [30] L. A. Zadeh, Similarity relations and fuzzy orderings, *Information Sciences* 3 (2) (1971) 177–200.

- [31] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning—I, *Information Sciences* 8 (3) (1975) 199–249.
- [32] W. Zeng, H. Li, Relationship between similarity measure and entropy of interval valued fuzzy sets, *Fuzzy Sets and Systems* 157 (11) (2006) 1477–1484.
- [33] H. Zhang, W. Zhang, C. Mei, Entropy of interval-valued fuzzy sets based on distance and its relationship with similarity measure, *Knowledge-Based Systems* 22 (6) (2009) 449–454.
- [34] P. Żywica, Similarity measures of Interval-Valued Fuzzy Sets in classification of uncertain data. Applications in the diagnosis of ovarian tumors, Ph.D. thesis, Adam Mickiewicz University, Poznań, Poland, (in Polish) (2016).
- [35] P. Żywica, Modelling medical uncertainties with use of fuzzy sets and their extensions, in: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Springer, Cham, 2018, pp. 369–380.
- [36] P. Żywica, A. Stachowiak, M. Wygralak, An Algorithmic Study of Relative Cardinalities for Interval-Valued Fuzzy Sets, *Fuzzy Sets and Systems* 294 (2016) 105–124.