Title: An effective similarity measurement under epistemic uncertainty

Author: Patryk Żywica, Michał Baczyński

Citation style: Żywica Patryk, Baczyński Michał. (2022). An effective similarity measurement under epistemic uncertainty. "Fuzzy Sets and Systems" Vol. 431 (2022), s. 160-177, doi:10.1016/j.fss.2021.02.013

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 431 (2022) 160–177

FUZZY sets and systems

www.elsevier.com/locate/fss

An effective similarity measurement under epistemic uncertainty

Patryk Żywica [a,*], Michał Baczyński [b]

[a] *Department of Artificial Intelligence, Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poznań, Uniwersytetu Poznańskiego 4, 61-614 Poznań, Poland*

[b] *Faculty of Science and Technology, University of Silesia in Katowice, Bankowa 14, 40-007 Katowice, Poland*

Received 7 February 2020; received in revised form 16 February 2021; accepted 18 February 2021
Available online 24 February 2021

Abstract

The epistemic uncertainty stems from the lack of knowledge and it can be reduced when the knowledge increases. Such interpretation works well with data represented as a set of possible states and therefore, multivalued similarity measures. Unfortunately, set-valued extensions of similarity measures are not computationally feasible even when the data is finite. Measures with properties that allow efficient calculation of their extensions, need to be found. Analysis of various similarity measures indicated logic-based (additive) measures as an excellent candidate. Their unique properties are discussed and efficient algorithms for computing set-valued extensions are given. The work presents results related to various classes of fuzzy set families: general ones, intervals of fuzzy sets, and their finite sums. The first case is related to the concept of the Fuzzy Membership Function Family, the second corresponds to the Interval-Valued Fuzzy Sets, while the third class is equivalent to the concept of Typical Interval-Valued Hesitant Fuzzy Sets.

Keywords: Similarity measure; Epistemic uncertainty; Set-valued extension; Interval-Valued Fuzzy Sets; IVFS; Hesitant Fuzzy Sets; HFS

1. Introduction

Any information can be presented as a set of possible states, one of which is the "true" or "real" one, not known at the moment due to the lack of knowledge. Thus, it is a way to describe or represent some uncertain information. The uncertainty is not of the probabilistic type, but stems from the lack of knowledge; it can be reduced when the knowledge increases. Such an interpretation of a set has an epistemic (conjunctive) nature – contrary to the ontic (disjunctive) one, when the set is understood as a complex, but certain, information [11,17]. We will thus denote **A** as a set of possible states that it represents. We clearly distinguish the description of an (unknown) object from the object itself (represented by one of the possible states). The notion of *uncertainty–aware* similarity measure that emphasizes this distinction was the primary motivation for this study. Let us assume, for example, that we want to compare two identical intervals: $[0.1, 0.8]$ and $[0.1, 0.8]$. We notice the total similarity of their description; however,

* Corresponding author.
E-mail addresses: bikol@amu.edu.pl (P. Żywica), michal.baczynski@us.edu.pl (M. Baczyński).

https://doi.org/10.1016/j.fss.2021.02.013

it does not imply a total similarity of the objects that are being described. This situation needs to be handled correctly by a similarity measure, but at the same time, it may imply significant computational difficulties.

Because of the many different approaches to data uncertainty, defining the similarity of epistemic data is a complex problem. Even the basic properties of the concept of similarity have not been clearly established and widely accepted. 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, 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 utilizing intervals or subsets, which results in multivalued similarity measures.

In general, set-valued extensions of similarity measures [9] are not computationally effective even when the data is finite. Even in the simplest case, the measure value is not obtained for a simple combination of the input fuzzy sets. This creates the need for further research into the problem, with particular regard to the computational aspects. We need to look for similarity measures with properties that allow efficient calculation of their extensions. Analysis of various families of similarity measures [15] indicates that logic-based measures are an excellent candidate for further research. A key feature of this family is the use of an aggregation operator to combine partial similarity degrees. This allows to change the scope of optimization and lead to very promising computational results.

The presented results are related to various classes of $\mathcal{F}(U)$ subsets:

- General families of fuzzy sets. In this case, $\mathbf{A}$ consists of any collection of individual membership functions. This approach is related to the concept of Fuzzy Membership Function Family introduced in [43].
- Intervals of fuzzy sets and their finite unions. The notion of *interval of fuzzy sets* was introduced in [29]. There is a one-to-one correspondence between intervals of fuzzy sets and Interval-Valued Fuzzy Sets (IVFS, [39]). However, such an epistemic interpretation of IVFS as a collection of fuzzy sets is not widely used.
- Typical Interval-Valued Hesitant Fuzzy Sets (TIVHFS, [7,34]).

Section 2 defines the essential concepts used in the rest of the work. The third section shows the motivation behind our research, which is the *uncertainty-aware* similarity measure. It summarizes some definitions and properties of such measures and relates them to the concept of a set-valued extension. Section 4 discusses the problem of the effective computing of such extensions of similarity measures with regard to the introduced classes of $\mathcal{F}(U)$ subsets. Examples of specific measures and discussion on the algorithmic approach to determining their values are also given. The fifth section contains the main results and deals with the most promising case of logic-based similarity measures. Their unique properties are discussed, and practical algorithms for computing set-valued extensions are given.

2. Definitions

Let $U = \{u_1, u_2, \ldots, u_n\}$ be a crisp universal set. A mapping $A\colon U \to [0,1]$ 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 $\mathbb{1}_X$. We say that fuzzy set A is a subset of fuzzy set B ($A \subset B$) if $A(u_i) \leq B(u_i)$ for all $u_i \in U$. Let $\mathcal{F}(U)$ be the family of all fuzzy sets in U, $\mathbf{A}(u_i)$ denote its subset and

$$\mathbf{A}(u_i) = \{x \in [0,1]\colon x = A(u_i), \text{ for some } A \in \mathbf{A}\}. \tag{1}$$

A binary operation $t\colon [0,1] \times [0,1] \to [0,1]$ 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 the minimum $t_{\min}(x,y) = \min(x,y)$, the product $t_{\text{prod}}(x,y) = xy$, and the Łukasiewicz $t_{\text{Łuk}}(x,y) = \max(0, x+y-1)$. A thorough investigation on t-norms can be found in the classical monograph of Klement et al. [24].

A function $n\colon [0,1] \to [0,1]$ is called a fuzzy negation [1] if it is decreasing, $n(0) = 1$ and $n(1) = 0$. The complement of fuzzy set A with respect to negation n is defined as a fuzzy set A' such that $A'(u_i) = n(A(u_i))$ for all $u_i \in U$. The most natural negation is defined as $n(x) = 1 - x$. Unless otherwise mentioned, it should be assumed that this negation was used.

Let $E \subset \mathcal{F}(U) \times \mathcal{F}(U)$ be such that

(S1) $(A,B) \in E$ if and only if $(B,A) \in E$,
(S2) if $(A,B) \in E$, then $(A, \mathbb{1}_U) \in E$.

Definition 1. A similarity measure of fuzzy sets is defined as a function $s\colon E \to [0,1]$ such that

(T1) for each $(A, B) \in E$, we have $s(A, B) = s(B, A)$,
(T2) for each $(A, D) \in E$ and $(B, C) \in E$ such that $A \subset B \subset C \subset D$ we have

$$s(A, D) \leq s(B, C), \tag{2}$$

(T3) for each $X \subset U$ such that $(\mathbb{1}_X, \mathbb{1}_{X^c}) \in E$ we have $s(\mathbb{1}_X, \mathbb{1}_{X^c}) = 0$ and $s(\mathbb{1}_X, \mathbb{1}_X) = 1$.

This definition coincides with the classical one proposed by Xuecheng [37] (see also [12,15,38]). The higher measure values indicate higher similarity of its arguments. It is usually assumed that $E = \mathcal{F}(U) \times \mathcal{F}(U)$, meaning that all fuzzy sets are comparable by a given similarity measure. However, some similarity measures (such as Jaccard index) can not be formally defined over the whole Cartesian product $\mathcal{F}(U) \times \mathcal{F}(U)$.

It is important to note that any fuzzy subset A of a finite universe U can be identified with a tuple $\big(A(u_1), \cdots, A(u_n)\big) \in [0,1]^n$ and therefore $\mathcal{F}(U)$ can be identified with the Cartesian product $[0,1]^n$.

Definition 2. Fuzzy set similarity measure $s\colon E \to [0,1]$ is called convex and continuous if E is convex, meaning that the set

$$X = \big\{(\mathbf{x}_A, \mathbf{x}_B) \in [0,1]^{2n} \colon (A, B) \in E\big\}, \tag{3}$$

is convex and the function $f\colon X \to [0,1]$ defined as

$$f(\mathbf{x}_A, \mathbf{x}_B) = s(A, B), \tag{4}$$

is continuous in the whole domain.

Definition 3 *([2]).* An n-argument aggregation operator is a mapping $\mathrm{Agg}\colon [0,1]^n \to [0,1]$ with the following properties:

1. if $x_i \leq y_i$ for all $i \in 1, \cdots, n$, then $\mathrm{Agg}(x_1, \cdots, x_n) \leq \mathrm{Agg}(y_1, \cdots, y_n)$,
2. $\mathrm{Agg}(1, \cdots, 1) = 1$,
3. $\mathrm{Agg}(0, \cdots, 0) = 0$.

If aggregation covers some indexed values x_i, it can be more convenient to use the following notation

$$\mathrm{Agg}(x_1, \cdots, x_i, \cdots, x_n) = \bigoplus_{i=1}^{n} x_i. \tag{5}$$

Definition 4. Median aggregation operator $\mathcal{M}\colon [0,1]^n \to [0,1]$ is defined as

$$M(x_1, \cdots, x_n) = \begin{cases} \frac{1}{2}\left(x_{(k)} + x_{(k+1)}\right), & \text{if } n = 2k \text{ is even,} \\ x_{(k)}, & \text{if } n = 2k-1 \text{ is odd,} \end{cases} \tag{6}$$

where $x_{(k)}$ is the k-th largest among $x_1, \cdots, x_n$.

3. Motivation

We have conducted an extensive analysis of many different approaches to defining similarity measures and related concepts (inclusion, subsethood, distance, dissimilarity, and entropy measures). The focus was put mainly on the properties of such measures regarding data uncertainty and their impact on the possibility of expressing real-world requirements and semantics of similarity. Thanks to the comparison of the properties of various similarity measures [6,8,12,13,21,25,26,28,32,33,37,40–42], it was possible to propose a concept of *uncertainty–aware* similarity measure [44].

Let $\mathbf{E} \subset \mathcal{P}\left(\mathcal{F}\left(\mathcal{U}\right)\right) \times \mathcal{P}\left(\mathcal{F}\left(\mathcal{U}\right)\right)$ be such that:

(E1) $(\mathbf{A}, \mathbf{B}) \in \mathbf{E}$ if and only if $(\mathbf{B}, \mathbf{A}) \in \mathbf{E}$,
(E2) $(\mathbf{A}, \mathbf{B}) \in \mathbf{E}$ if $\left(\mathbf{A}, \{\mathbb{1}_U\}\right) \in \mathbf{E}$,
(E3) $(\mathbf{A}, \mathbf{B}) \in \mathbf{E}$ if and only if for any fuzzy sets $A \in \mathbf{A}$, $B \in \mathbf{B}$:

$$\left(\{A\}, \{B\}\right) \in \mathbf{E}. \tag{7}$$

Definition 5. A function $\widetilde{s}\colon \mathbf{E} \to \mathcal{P}([0,1])$ is an *uncertainty–aware* similarity measure if it satisfies following conditions:

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

$$\widetilde{s}(\mathbf{A}, \mathbf{B}) = \widetilde{s}(\mathbf{B}, \mathbf{A}). \tag{8}$$

(P2) If $(\mathcal{F}(U), \mathcal{F}(U)) \in \mathbf{E}$ then

$$\widetilde{s}(\mathcal{F}(U), \mathcal{F}(U)) = [0,1]. \tag{9}$$

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

$$1 \in \widetilde{s}(\mathbf{A}, \mathbf{B}), \tag{10}$$

$$0 \in \widetilde{s}(\mathbf{A}, \mathbf{C}). \tag{11}$$

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

$$\widetilde{s}(\{A\}, \{B\}) = \{a\}, \quad \text{for some } a \in [0,1]. \tag{12}$$

(P5) For any $(\mathbf{A}, \mathbf{C}) \in \mathbf{E}$, $(\mathbf{B}, \mathbf{D}) \in \mathbf{E}$ such that $\mathbf{A} \subset \mathbf{B}$ and $\mathbf{C} \subset \mathbf{D}$,

$$\widetilde{s}(\mathbf{A}, \mathbf{C}) \subset \widetilde{s}(\mathbf{B}, \mathbf{D}). \tag{13}$$

(P6) For any $(\mathbf{A}, \mathbf{D}) \in \mathbf{E}$ and $(\mathbf{B}, \mathbf{C}) \in \mathbf{E}$ and for all $A \in \mathbf{A}$, $B \in \mathbf{B}$, $C \in \mathbf{C}$, $D \in \mathbf{D}$ such that $A \subset B \subset C \subset D$ we have

$$s_{ad} \leq s_{bc}, \tag{14}$$

where

$$\widetilde{s}\big(\{A\}, \{D\}\big) = \{s_{ad}\} \text{ and } \widetilde{s}\big(\{B\}, \{C\}\big) = \{s_{bc}\}. \tag{15}$$

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 an unknown object, the similarity should also remain unknown. On the other hand, if the information is complete (is reduced 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 pieces of epistemic information could be similar to a degree 1 only if they share at least one common state (10). On the other hand, if they are inconsistent, then 0 should be a possible value of their similarity (11). In general, when a degree of uncertainty decreases so does similarity measure - which has been captured in (P5). Consequently, for any pair of possible states, their similarity measure belongs to the similarity of any uncertain information that they belong to. Finally, property (P6) imposes the monotonicity with respect to fuzzy set inclusion. It should be noted that inclusion relation plays a purely technical role in this formula – it only guarantees the proper ordering of the membership functions. For this reason, a deeper interpretation or generalization does not always make sense.

This definition captures the ideas presented in many previous works [30,31,36,45]. Moreover, it turns out that this definition is consistent with the more general concept of set-valued extension considered by Couso and Bustince [9] in which the uncertain input, treated as a set of states, is used to calculate its image under the original fuzzy set similarity measure. This approach is formalized in the following definition.

Definition 6 *(Set-valued extension, [9]).* Let $s\colon E \to [0,1]$ be a similarity measure of fuzzy sets. Function $[s]:\mathbf{E}\to \mathcal{P}([0,1])$ can be defined in the following way:

$$[s](\mathbf{A},\mathbf{B}) = \{s(A,B) : A\in\mathbf{A}, B\in\mathbf{B}\}, \tag{16}$$

where

$$\mathbf{E} = \{(\mathbf{A},\mathbf{B})\in\mathcal{P}(\mathcal{F}(\mathcal{U}))\times\mathcal{P}(\mathcal{F}(\mathcal{U})) : \mathbf{A}\times\mathbf{B}\subset E\}. \tag{17}$$

Theorem 1. *For each convex and continuous fuzzy set similarity measure $s\colon E\to[0,1]$, its set-valued extension $[s]$ is an* uncertainty–aware *similarity measure.*

Proof. Let $s : E\to[0,1]$ satisfy the assumptions. Let $[s]:\mathbf{E}\to\mathcal{P}([0,1])$ be defined according to the Definition 6. The first step is to check whether $\mathbf{E}$ satisfies conditions (E1)-(E3) required before definition of an *uncertainty aware* similarity measure. Two first properties follow directly from the definition. Let check the third condition,

$$\begin{aligned}(\mathbf{A},\mathbf{B})\in\mathbf{E} &\overset{(17)}{\iff} \mathbf{A}\times\mathbf{B}\subset E \iff \forall_{\substack{A\in\mathbf{A}\\B\in\mathbf{B}}}(A,B)\in E\\ &\iff \forall_{\substack{A\in\mathbf{A}\\B\in\mathbf{B}}}\{A\}\times\{B\}\subset E \overset{(17)}{\iff} \forall_{\substack{A\in\mathbf{A}\\B\in\mathbf{B}}}\big(\{A\},\{B\}\big)\in\mathbf{E}.\end{aligned} \tag{18}$$

Now we will prove each property separately:

(P1) The symmetry of the resulting measure is due to (T1).

(P2) Let $(\mathcal{F}(U),\mathcal{F}(U))\in\mathbf{E}$. Then $(\mathbb{1}_U,\mathbb{1}_\emptyset),(\mathbb{1}_U,\mathbb{1}_U)\in E$. Moreover, from (T3) we know that $s(\mathbb{1}_U,\mathbb{1}_U)=1$ and $s(\mathbb{1}_U,\mathbb{1}_\emptyset)=0$. Let for any $\alpha\in[0,1]$ define set A_α such that $A_\alpha(u_i)=\alpha$ ($A_0=\mathbb{1}_\emptyset$ and $A_1=\mathbb{1}_U$). Because E is convex we know that $(\mathbb{1}_U,A_\alpha)\in E$ for any $\alpha\in[0,1]$. Moreover, function $s':[0,1]\to[0,1]$ defined as $s'(\alpha)=s(\mathbb{1}_U,A_\alpha)$ is continuous real function such that $s'(0)=0$ and $s'(1)=1$, which by Intermediate value theorem proves that $[s](\mathcal{F}(U),\mathcal{F}(U))=[0,1]$.

(P3) This follows directly from (T3) similarly as in the proof of (P2).

(P4) This property follows directly from (16). Let $A,B\in\mathcal{F}(U)$ such that $(\{A\},\{B\})\in\mathbf{E}$, then

$$[s](\{A\},\{B\}) = \{s(A,B) : A\in\{A\}, B\in\{B\}\} = \{s(A,B)\}. \tag{19}$$

(P5) Let $(\mathbf{A},\mathbf{C})\in\mathbf{E}$, $(\mathbf{B},\mathbf{D})\in\mathbf{E}$ be such that $\mathbf{A}\subset\mathbf{B}$ and $\mathbf{C}\subset\mathbf{D}$. We have

$$\begin{aligned}[s](\mathbf{A},\mathbf{C}) &= \{s(A,C) : A\in\mathbf{A}, C\in\mathbf{C}\}\\ &\subset \{s(A,C) : B\in\mathbf{B}, D\in\mathbf{D}\} = [s](\mathbf{B},\mathbf{D}).\end{aligned} \tag{20}$$

(P6) Let $(\mathbf{A},\mathbf{D}),(\mathbf{B},\mathbf{C})\in\mathbf{E}$ and $A\in\mathbf{A}$, $B\in\mathbf{B}$, $C\in\mathbf{C}$, $D\in\mathbf{D}$ be such that $A\subset B\subset C\subset D$. From (16) we may observe that

$$\{s(A,D)\} = [s]\big(\{A\},\{D\}\big) = \{s_{ad}\}, \tag{21}$$

$$\{s(B,C)\} = [s]\big(\{B\},\{C\}\big) = \{s_{bc}\}. \tag{22}$$

Thanks to (T2) we know that

$$s_{ad} = s(A,D)\le s(B,C) = s_{bc}, \tag{23}$$

which completes the proof. □

An additional result in opposite direction may be derived from (P4) and (P5) properties.

Remark 1. Any uncertainty-aware similarity measure $\widetilde{s}$ derives a similarity measure s for fuzzy sets, such that the images of $\widetilde{s}$ for any pair $(\mathbf{A},\mathbf{B})$ include the corresponding set of images of s, i.e.:

$$\widetilde{s}(\mathbf{A},\mathbf{B})\supseteq\{s(A,B) : A\in\mathbf{A}, B\in\mathbf{B}\}. \tag{24}$$

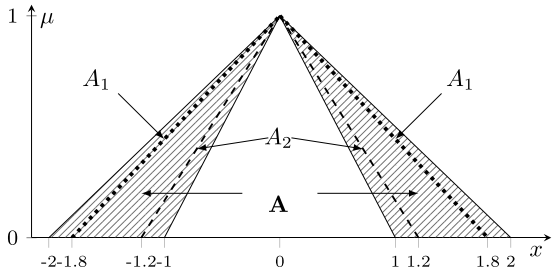

Fig. 1. Visualization of the Interval–Valued Fuzzy Set $\hat{A}$ and *interval of fuzzy sets* associated with it (shaded area). Dashed lines represent two embedded fuzzy sets $A_1 \in \hat{A}$ and $A_2 \in \hat{A}$.

Hence, the axiomatic approach presented in the uncertainty-aware similarity measure and the constructive one with the set-valued extension lead to similar results, which is consistent with the general theoretical results [10].

It is important to note that those concepts are not equivalent. There exist many *uncertainty-aware* similarity measures that do not fall under set-valued extension scheme. Consider the following example:

$$\widetilde{s}(\mathbf{A},\mathbf{B}) = \begin{cases} \{0\}, & \text{if } \mathbf{A}=\{A\}, \mathbf{B}=\{B\} \text{ and } A \neq B, \\ \{1\}, & \text{if } \mathbf{A}=\{A\}, \mathbf{B}=\{B\} \text{ and } A = B, \\ [0,1], & \text{otherwise.} \end{cases} \tag{25}$$

4. Computational study on set-valued extensions of similarity measures

In real-life scenarios, most of the time, the data is finite in terms of both attributes and records. It is particularly true in the case of information obtained directly from a human. Moreover, computers can only manage finite data. Having this in mind, in this paper, we correspond to the case where data can be reliably represented in a finite way.

In general, set-valued extensions of similarity measures are not computationally effective even while data is finite. Similarity measures are not monotonically increasing and decreasing with respect to some combination of components, neither satisfy linearity conditions (see [9]). For this reason, even in the simplest case of the interval-valued extension, the measure value bounds are not obtained for a simple combination of the input family of fuzzy sets bounds.

In this paper, we distinguish 4 cases:

1. General family of fuzzy sets (FoFS).
   In this case, the only restriction is that $\mathbf{A}$ is a closed, nonempty subset of $\mathcal{F}(U)$; hence it consists of any collection of individual membership functions. This approach is analogous to the concept of Fuzzy Membership Function Family introduced in [43].
2. Interval of fuzzy sets (IoFS).
   The notion of *interval of fuzzy sets* was introduced in [29]. It is defined as

   $$\mathbf{A} = [\underline{A}, \overline{A}] = \left\{ A \in \mathcal{F}(U) : \underline{A} \subseteq A \subseteq \overline{A} \right\}. \tag{26}$$

   There is a one-to-one correspondence between intervals of fuzzy sets and Interval-Valued Fuzzy Sets (see Fig. 1). The difference between those two concepts is in their interpretation. IVFS are understood as generalized fuzzy sets where membership values are sub-intervals of $[0, 1]$ instead of a single number. On the other hand, intervals of fuzzy sets are interpreted as collections of fuzzy sets between the lower and upper bound. Because the epistemic interpretation of IVFS as a collection of fuzzy sets is not widely accepted, we will use the notion of the interval of fuzzy sets throughout the rest of this work.
3. Union of fuzzy set intervals (UIoFS).
   This case covers any finite union of fuzzy set intervals, hence this allows to introduce discontinuities into the membership values:

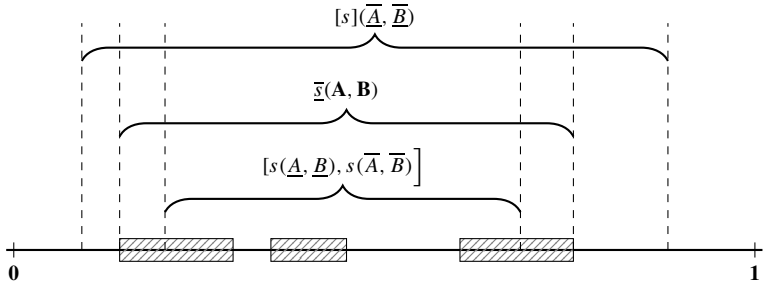

Fig. 2. Relations between different approximations of $[s](\mathbf{A}, \mathbf{B})$ (dashed rectangles).

$$\mathbf{A} = \bigcup_{1 \le i \le k} \left[\underline{A}_i, \overline{A}_i\right] = \left\{A \in \mathcal{F}(U) : \exists_{1 \le i \le k} \quad \underline{A}_i \subseteq A \subseteq \overline{A}_i\right\}. \tag{27}$$

4. Typical interval-valued hesitant fuzzy sets (TIVHFS).
   In this case the information can be characterized in the sense of possible membership values for given element of the universe:

$$\mathbf{A}(u_j) = \bigcup_{1 \le j \le k_j} \left[\underline{a}_i^j, \overline{a}_i^j\right], \tag{28}$$

   for some $0 \le \underline{a}_i^j \le \overline{a}_i^j \le 1$.

In the following, n will denote the size of the universal set $n = |U|$, and k the number of intervals of fuzzy sets or disjoint intervals that are allowed as a membership value of any $A \in \mathbf{A}$. To simplify the worst-case analysis, without loss of generality, we assume that each membership value belongs to the union of exactly k intervals. For the first case, to make the results comparable, we assume that the number of membership functions is of the $O(k^n)$ order.

For the first case (FoFS), the naive approach of iterating through all states requires $O(k^n)$ calculations of original similarity measure (which most often requires $O(n)$ operations itself). This case is very complex, and one may only try to approximate the final similarity value, which may be an arbitrary nonempty subset of $[0, 1]$.

The naive algorithm for the second case (IoFS) is, in some sense, even more complex. Although the data representation is finite, the set-valued extension requires an infinite number of values to be checked. Hence, two naive approaches are possible. The first is based on reducing it to the previous case by replacing the interval of fuzzy sets with a finite number of membership functions. The second is to use general optimization algorithms, such as Brent algorithm [5]. Both solutions are very inefficient in terms of computation complexity and do not guarantee an accurate result. Particular problems in this class have been the subject of previous research [23,27,45,46], and algorithms (exact and approximate) for many functions of fuzzy sets including similarity measures are known.

The case of UIoFS is a straightforward generalization of the previous one. Having any solution for IoFS case with $O(f(n))$ complexity we can easily construct an algorithm that will solve UIoFS case in $O(kf(n))$ by iterating through all k intervals and merging the results.

In the fourth case (TIVHFS), if there is known computation method for the second case with a complexity of $O(f(n))$, then the naive algorithm requires $O(k^n f(n))$ operations.

When it is not possible to compute a set-valued extension of the similarity measure effectively, we will try to provide a fast algorithm that finds the smallest interval that contains the actual value (bounding interval), denoted by $\underline{s}(\mathbf{A}, \mathbf{B}) = [\underline{\underline{s}}(\mathbf{A}, \mathbf{B}), \overline{\underline{s}}(\mathbf{A}, \mathbf{B})]$. This value corresponds to the second type of extensions proposed in [9] – the min/max extension.

It is important to note that a simple approach to finding bounding intervals via substituting the union of intervals by a single interval that covers all of them, will potentially yield a too wide result – denoted by $[s](\underline{A}, \overline{B})$. On the other hand, computing similarity only for extreme memberships (or any other selected representatives) – $\underline{A}, \overline{A} \in \mathbf{A}$ and $\underline{B}, \overline{B} \in \mathbf{B}$ – will yield a too narrow result. Following inclusions hold (see Fig. 2):

$$\left[s(\underline{A}, \underline{B}), s(\overline{A}, \overline{B})\right] \subseteq \underline{s}(\mathbf{A}, \mathbf{B}) \subseteq [s](\underline{A}, \overline{B}) \text{ and } [s](\mathbf{A}, \mathbf{B}) \subseteq \underline{s}(\mathbf{A}, \mathbf{B}). \tag{29}$$

Table 1
Computational complexity of set-valued extensions of similarity measures. If lower and upper bounds differ in complexity, worst case is given.

	Case 1 (FoFS)	Case 2 (IoFS)	Case 3 (UIoFS)	Case 4 (TIVHFS)
Minkowski distance				
exact value	$O(nk^n)$	$O(n)$	$O(nk)$	$O(nk^n)$
bounding interval	–	–	–	$O(nk\log k)$
Jaccard index				
exact value	$O(nk^n)$	$O(n\log n)$	$O(nk\log n)$	$O(nk^n\log n)$
bounding interval	–	–	–	$O(nk+n\log n)$
Simple matching coefficient				
exact value	$O(nk^n)$	$O(n)$	$O(nk)$	$O(nk^n)$
bounding interval	–	–	–	$O(nk\log k)$
Logic-based				
exact				
– mean aggregation	$O(nk^n)$	$O(n)$	$O(nk)$	$O(nk^n)$
– median aggreg.	$O(nk^n)$	$O(n)$	$O(nk)$	$O(n^3k^3)$
– min/max aggreg.	$O(nk^n)$	$O(n)$	$O(nk)$	$O(n^2k^2)$
bounding interval	–	–	–	$O(nk\log k)$

A deeper formal discussion on this topic can be found in [9,10].

After a general discussion on computational problems related to the set-valued extensions of similarity measures, we will show examples of extensions of well-known measures. Using the classification proposed by Cross and Sudkamp [15], the examples of distance and set-theory based similarity measures will be discussed in the following Subsections. Section 5 deals with the computation of the set-valued extensions of the logic-based similarity measures, which is the main result presented in this work.

4.1. Set-valued extensions of similarity measures

The following section discusses the set-valued extensions of known similarity measures obtained using Definition 6. As will be shown further in this section, the computation of some similarity measures is difficult, while other measures can be calculated using simple formulas. The most important facts from this subsection were gathered in Table 1.

4.1.1. Distance based similarity measures

The Minkowski distance is very often used as a measure of similarity

$$s_{d_r}(A,B)=1-\frac{d_r(A,B)}{\sqrt[r]{|U|}}=1-\frac{1}{\sqrt[r]{|U|}}\sqrt[r]{\sum_{u_i\in U}\left|A(u_i)-B(u_i)\right|^r}. \tag{30}$$

It is defined over the set $E=E_U=\mathcal{F}(U)\times\mathcal{F}(U)$. Such measure meets the assumptions of Theorem 1. Thanks to this, its set-valued extension is an *uncertainty–aware* similarity measure.

There are well-known formulas for calculating the set-valued extension of the Minkowski distance in the IoFS case [45]. In the TIVHFS case, computing the exact value requires naive iteration through all combinations of intervals. On the other hand, the bounding interval $\underline{\overline{s}}(\mathbf{A},\mathbf{B})$ can be computed effectively. Lower bound of $\underline{\overline{s}}(\mathbf{A},\mathbf{B})$ can be calculated in $O(n)$ using

$$\underline{\overline{s}}(\mathbf{A},\mathbf{B})=1-\frac{1}{|U|}\left(\sum_{u\in U}\max\left\{|\underline{A}(u)-\overline{B}(u)|,|\overline{A}(u)-\underline{B}(u)|\right\}^r\right)^{\frac{1}{r}}, \tag{31}$$

where $\underline{A}, \overline{A}, \underline{B}, \overline{B}$ represent lower and upper bounds in $\mathbf{A}$ and $\mathbf{B}$. Upper bound can be computed in $O(nk \log k)$ using the Bentley–Ottmann algorithm [4] to find the two closest points in $\mathbf{A}(u_i)$ and $\mathbf{B}(u_i)$ for each $u_i \in U$.

4.2. Set-theory-based similarity measures

4.2.1. Jaccard index

The Jaccard index is the most commonly used similarity measure. It formalizes the observation that for any 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

$$s_J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \text{ where } |A \cup B| \neq 0. \tag{32}$$

Because $A \cap B \subset A \cup B$, the 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. Unfortunately, the unambiguous definition of the similarity value in the case when $|A \cup B| = 0$ is not possible. Thus, the domain of similarity measure is defined as

$$E_J = \mathcal{F}(U) \times \mathcal{F}(U) \setminus \left\{(\mathbb{1}_\emptyset, \mathbb{1}_\emptyset)\right\} = E_\emptyset. \tag{33}$$

The Jaccard index was successfully extended to the IoFS case first by Nguyen and Kreinovich [30], then was generalized to solve other related problems [42,45,46]. Proposed algorithms allow for an efficient calculation of the lower bound in $O(n \log n)$ and upper in $O(n)$ operations. Those algorithms can be directly extended to compute bounding intervals in TIVHFS case, which will result in a complexity of $O(nk + n \log n)$ and $O(nk \log k)$, respectively. Exact value in the fourth case (TIVHFS) still requires naive iteration throughout all interval combinations.

4.2.2. Simple matching coefficient

One of the frequent critics against the Jaccard's index is that it is not taking into account common deficiencies in the two sets. This issue is solved by the Simple matching coefficient:

$$s_{\text{smc}}(A, B) = \frac{|A \cap B| + |A' \cap B'|}{|U|}. \tag{34}$$

The following transformation

$$\begin{aligned} s_{\text{smc}}(A, B) &= \frac{1}{|U|} \sum_{u_i \in U} \min(A(u_i), B(u_i)) + \min(1 - A(u_i), 1 - B(u_i)) \\ &= 1 - \frac{1}{|U|} \sum_{u_i \in U} |A(u_i) - B(u_i)|, \end{aligned} \tag{35}$$

shows that SMC is a special case of the Minkowski distance based similarity measure for $r = 1$ (see Subsection 4.1.1). It is interesting to note that, on the one hand, Simple matching coefficient can be treated as a generalization of the Jaccard index, and on the other we get a simpler algorithm to calculate it.

5. Logic-based similarity measures

Logic-based measures [15,19,20] 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 inclusion and similarity measures. In classical 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 SN–implications, R–implications and QL–implications [1,35].

5.1. Definition and properties

Definition 7. Logic-based similarity measure is defined as an aggregation of the equality values $\Psi : G \to [0,1]$ obtained for all elements of the universe

$$s_\Psi(A,B) = \bigoplus_{u\in U} \Psi(\mu_A(u), \mu_B(u)). \tag{36}$$

The domain of such a function is defined as

$$E_G = \left\{ (A,B) \in \mathcal{F}(U)^2 : \forall_{u_i \in U} \ (A(u_i), B(u_i)) \in G \right\}, \tag{37}$$

where $G \subset [0,1]^2$ and

$(G1)$ $(a,b) \in G$ if and only if $(b,a) \in G$,
$(G2)$ $(a,b) \in G$ if $(a,1) \in G$.

This definition generalizes the concept of *additive similarity measure* recently proposed by Couso [13,14]. It should be noted that replacing arithmetic mean by any other aggregation operator does not affect any of the assumptions or properties. Moreover, this approach allows for direct integration of the weights of individual elements of the universe into a similarity measure.

Theorem 2 *(Couso & Sanchez, [13]). Function $s_\Psi : E_G \to [0,1]$ is a similarity measure when*

$(\Psi 1)$ *for any* $(a,b) \in G$

$$\Psi(a,b) = \Psi(b,a), \tag{38}$$

$(\Psi 2)$ *for any* $(a,d),(b,c) \in G$ *such that* $a \le b \le c \le d$ *we have that*

$$\Psi(a,d) \le \Psi(b,c), \tag{39}$$

$(\Psi 3)$ $\Psi(0,1) = 0$, $\Psi(1,1) = 1$ *and* $\Psi(0,0) = 1$.

Remark 2. Similarity measure $s_\Psi : E_G \to [0,1]$ meets the assumptions of the Theorem 1 if $\Psi : G \to [0,1]$ is continuous in the whole domain and G is convex.

Remark 3. All fuzzy equivalences defined by Fodor and Roubens [18] meet $(\Psi 1)$–$(\Psi 3)$.

Lemma 3. *Let G be a convex subset of $[0,1]^2$ which satisfy $(G1)$ and $(G2)$. For any implication operator I, continuous on G, co-implication operator (see [22]) $\Psi : G \to [0,1]$ given by*

$$\Psi(a,b) = \min\left(I(a,b), I(b,a)\right) \tag{40}$$

fulfills properties $(\Psi 1)$–$(\Psi 3)$.

Proof. $(\Psi 1)$ This is fulfilled for any implication I.
$(\Psi 2)$ Let assume that $(a,d),(b,c) \in G$ and $a \le b \le c \le d$. From basic properties of implication operator we know that when $x \le y$

$$I(x,y) \ge I(y,x). \tag{41}$$

This allows us to observe that

$$\Psi(a,d) = I(d,a) \text{ and } \Psi(b,c) = I(c,b). \tag{42}$$

Now we only need to observe that again due to monotonicity of I

$$\Psi(a,d) = I(d,a) \le I(c,a) \le I(c,b) = \Psi(b,c). \tag{43}$$

Table 2
Co-implication operators obtained from nine basic implication operators.

Implication name	$I(x,y)$	$\Psi(x,y)$
Łukasiewicz, $I_{\text{ŁK}}$	$\min(1, 1-x+y)$	$1-\lvert x-y\rvert$
Gödel, I_{GD}	$\begin{cases}1, & x \le y\\ y, & x > y\end{cases}$	$\begin{cases}1, & x = y\\ \min(x,y), & x \ne y\end{cases}$
Reichenbach, I_{RC}	$1-x+xy$	$1-\max(x,y)+xy$
Kleene-Dines, I_{KD}	$\max(1-x, y)$	$\begin{cases}\min(x,y), & 1-x \le y\\ \min(1-x, 1-y), & 1-x > y\end{cases}$
Goguen, I_{GG}	$\begin{cases}1, & x \le y\\ \frac{y}{x}, & x > y\end{cases}$	$\begin{cases}1, & x = y = 0\\ \frac{\min(x,y)}{\max(x,y)}, & x \ne 0, y \ne 0\end{cases}$
Rescher, I_{RS}	$\begin{cases}1, & x \le y\\ 0, & x > y\end{cases}$	$\begin{cases}1, x = y\\ 0, & x \ne y\end{cases}$
Yager, I_{YG}	$\begin{cases}1, & x = y = 0\\ y^x, & x, y > 0\end{cases}$	$\begin{cases}1, & x = y = 0\\ \min(x^y, y^x), & x, y > 0\end{cases}$
Weber, I_{WB}	$\begin{cases}1, & x < 1\\ y, & x = 1\end{cases}$	$\begin{cases}1, & x, y < 1\\ \min(x,y), & x = 1 \text{ or } y = 1\end{cases}$
Fodor, I_{FD}	$\begin{cases}1, & x \le y\\ \max(1-x, y), & x > y\end{cases}$	$\begin{cases}1-x, & y \le \min(x, 1-x)\\ y, & 1-x < y < x\\ 1-y, & x < y < 1-x\\ x, & y > \max(x, 1-x)\end{cases}$

(Ψ3) Using basic properties of implication operator I we have that

$$\Psi(0,1) = \min(I(0,1),\ I(1,0)) = \min(I(0,1), 0) = 0, \tag{44}$$

$$\Psi(0,0) = I(0,0) = 1, \tag{45}$$

$$\Psi(1,1) = I(1,1) = 1. \quad \square \tag{46}$$

Lemma 4. *Let G be a convex subset of $[0,1]^2$ which satisfy $(G1)$, $(G2)$ and such that*

$$(x,y) \in G \text{ if and only if } (n(x), n(y)) \in G. \tag{47}$$

For any implication operator I, continuous on G, equality index (see [19,20]) $\Psi : G \to [0,1]$ given by

$$\Psi(a,b) = \frac{1}{2}\Big(\min\big(I(a,b), I(b,a)\big) + \min\big(I(n(a), n(b)), I(n(b), n(a))\big)\Big) \tag{48}$$

fulfills properties (Ψ1)–(Ψ3).

It is easy to observe that for SN–implications, co-implication operator and equality index are equivalent. Moreover, note that Ψ need not be defined in terms of co-implication nor equality index. This opens wide possibilities for research on the selection of appropriate comparison function. Table 2 and Fig. 3 presents basic implication operators along with their co-implication operator.

One may ask why both co-implication operators, as well as equality indexes, are based on minimum conjunction. In the case of fuzzy equivalences, it was proved that co-implication is the only possible formula [18].

Example 1. Co-implication for Reichenbach implication operator defined using product conjunction is the following:

$$\Psi(a,b) = I(a,b)I(b,a) = (1-x+xy)(1-y+xy). \tag{49}$$

This function fails to meet (Ψ3).

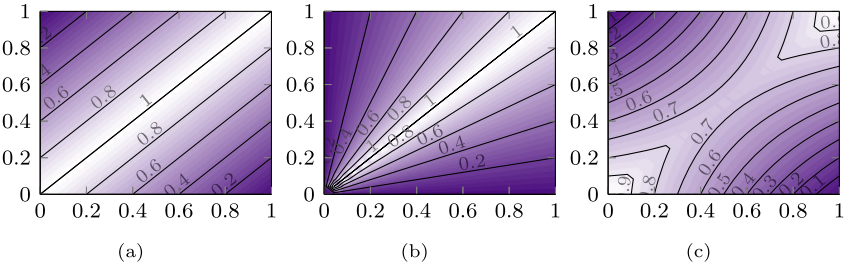

Fig. 3. Contour plots of co-implication operators obtained from Łukasiweicz (a), Goguen (b) and Reichenbach (c) implication operations.

Example 2. Co-implication for Łukasiewicz implication operator defined using product conjunction is following:

$$\Psi(a,b) = I(a,b)I(b,a) = \min(1, 1-x+y)\min(1, 1-y+x) = -|a-b|+1. \tag{50}$$

This function is equivalent to the operator defined using the minimum conjunction. This observation can be generalized to any R-implication and conjunction t-norm.

5.2. Computing the bounding interval of set-valued extension in the TIVHFS case

In this Section, we will first discuss the calculation of the bounding interval in TIVHFS case. The first important observation is changing the scope of optimization. Instead of optimizing the whole sum, we can optimize per element of the universe

$$\underline{\underline{s}}(\mathbf{A},\mathbf{B}) = \min_{\substack{A\in\mathbf{A}\\B\in\mathbf{B}}} \frac{1}{|U|} \bigoplus_{u_i\in U} \Psi(A(u_i), B(u_i)) = \frac{1}{|U|} \bigoplus_{u_i\in U} \min_{\substack{a\in\mathbf{A}(u_i)\\b\in\mathbf{B}(u_i)}} \Psi(a,b), \tag{51}$$

$$\overline{\underline{s}}(\mathbf{A},\mathbf{B}) = \max_{\substack{A\in\mathbf{A}\\B\in\mathbf{B}}} \frac{1}{|U|} \bigoplus_{u_i\in U} \Psi(A(u_i), B(u_i)) = \frac{1}{|U|} \bigoplus_{u_i\in U} \max_{\substack{a\in\mathbf{A}(u_i)\\b\in\mathbf{B}(u_i)}} \Psi(a,b). \tag{52}$$

Thanks to the property (Ψ2) we know that

$$\min_{\substack{a\in\mathbf{A}(u_i)\\b\in\mathbf{B}(u_i)}} \Psi(a,b) = \min\Big(\Psi\big(\underline{A}(u_i), \overline{B}(u_i)\big), \Psi\big(\overline{A}(u_i), \underline{B}(u_i)\big)\Big), \tag{53}$$

where $\underline{A}$, $\overline{A}$, $\underline{B}$, $\overline{B}$ represent lower and upper bounds in $\mathbf{A}$ and $\mathbf{B}$. This gives a simple formula for direct calculation of the lower bound. Due to the same property, the upper bound can be also easily computed, similarly as in the case of Minkowski distance, in $O(nk\log k)$ using the Bentley–Ottmann algorithm to find the two closest points in $\mathbf{A}(u_i)$ and $\mathbf{B}(u_i)$. This reasoning also shows that the exact value in the IoFS case can be computed in $O(n)$.

5.3. Computing the exact set-valued extension in TIVHFS case

In the following, we will show the general method for calculating the exact value for the wide family of logic-based similarity measures in the TIVHFS case. To make this possible, some reasonable assumptions must be made. First of all, Ψ have to be continuous in its domain. Moreover, for computational reasons, we will assume a finite numerical representation of $[0,1]$ interval using β different possible values. One may observe that the original calculations may be converted to the case when only $O(nk)$ unique values in $[0,1]$ are used as input. For this reason, we will assume that $\beta = O(nk)$.

Under those assumptions, we can transform the similarity measure in the following way

$$[s](\mathbf{A},\mathbf{B}) = \left\{ \bigoplus_{u_i\in U} \Psi(a_i,b_i) : a_i \in \mathbf{A}(u_i), b_i \in \mathbf{B}(u_i) \right\}$$

$$= \left\{ \bigoplus_{u_i \in U} \phi_i : \phi_i = \Psi(a_i, b_i), a_i \in \mathbf{A}(u_i), b_i \in \mathbf{B}(u_i) \right\}$$
$$= \left\{ \bigoplus_{u_i \in U} \phi_i : \phi_i \in \Phi_i \right\}, \tag{54}$$

where thanks to the continuity of Ψ

$$\Phi_i = \{\Psi(a_i, b_i) : a_i \in \mathbf{A}(u_i), b_i \in \mathbf{B}(u_i)\} = \bigcup_{1 \le j \le k} \left[\underline{\phi}_i^j, \overline{\phi}_i^j \right], \tag{55}$$

for some $\underline{\phi}_i^j$ and $\overline{\phi}_i^j$. Continuing (54) we have that

$$(54) = \left\{ \bigoplus_{u_i \in U} \phi_i : \phi_i \in \Phi_i \right\} = \left\{ \bigoplus_{u_i \in U} \phi_i : \phi_i \in \bigcup_{1 \le j \le k} \left[\underline{\phi}_i^j, \overline{\phi}_i^j \right] \right\}$$
$$= \bigcup_{\substack{(j_1, \cdots, j_n) \\ 1 \le j_i \le k}} \left\{ \bigoplus_{u_i \in U} \phi_i : \phi_i \in \left[\underline{\phi}_i^{j_i}, \overline{\phi}_i^{j_i} \right] \right\} \subseteq \bigcup_{\substack{(j_1, \cdots, j_n) \\ 1 \le j_i \le k}} \left[\bigoplus_{u_i \in U} \underline{\phi}_i^j, \bigoplus_{u_i \in U} \overline{\phi}_i^j \right]. \tag{56}$$

Now the question is when this set inclusion is actually an equality.

Definition 8. Aggregation operator $\bigoplus : [0, 1]^n \to [0, 1]$ is called *reducible to interval* if for any $0 \le \underline{\phi}_i \le \overline{\phi}_i \le 1$

$$\left\{ \bigoplus_{1 \le i \le n} \phi_i : \phi_i \in \left[\underline{\phi}_i, \overline{\phi}_i \right] \right\} = \left[\bigoplus_{1 \le i \le n} \underline{\phi}_i, \bigoplus_{1 \le i \le n} \overline{\phi}_i \right]. \tag{57}$$

It is easy to see that any additive and homogeneous aggregation operator is *reducible to interval*. As a result for all weighted means there is an equality in (56). Following lemma show that median is also *reducible to interval*. Similar reasoning can be also used for minimum and maximum aggregation operators.

Lemma 5. *Median aggregation operator* $\mathcal{M} : [0, 1]^n \to [0, 1]$ *is* reducible to interval.

Proof. We need to show that

$$L = \left[\bigoplus_{1 \le i \le n} \underline{\phi}_i, \bigoplus_{1 \le i \le n} \overline{\phi}_i \right] \subseteq \left\{ \bigoplus_{1 \le i \le n} \phi_i : \phi_i \in \left[\underline{\phi}_i, \overline{\phi}_i \right] \right\} = R. \tag{58}$$

Let assume that there exists $s \in L$ such that $s \notin R$. This means that

$$\bigoplus_{1 \le i \le n} \underline{\phi}_i \le s \le \bigoplus_{1 \le i \le n} \overline{\phi}_i, \tag{59}$$

and

$$s \ne \bigoplus_{1 \le i \le n} \phi_i \quad \text{for any} \quad \phi_i \in \left[\underline{\phi}_i, \overline{\phi}_i \right]. \tag{60}$$

We have two cases:

1. s is too big to be median of any sequence $(\phi_1 \cdots, \phi_n), \phi_i \in \left[\underline{\phi}_i, \overline{\phi}_i \right]$,
2. s is too small to be median of any sequence $(\phi_1 \cdots, \phi_n), \phi_i \in \left[\underline{\phi}_i, \overline{\phi}_i \right]$.

In the first case, for more than half of the indexes it must hold that $\overline{\phi}_i < s$. But this contradicts the assumption that

$$s \leq \bigoplus_{i=1}^{n} \overline{\phi}_i. \tag{61}$$

Similarly, in the second case, for more than half of the indexes it must hold that $\underline{\phi}_i > s$. This contradicts the assumption that

$$\bigoplus_{i=1}^{n} \underline{\phi}_i \leq s, \tag{62}$$

which completes the proof. □

Not all aggregation operators are *reducible to interval*. Consider following example

$$\mathrm{Agg}(x_1, \cdots, x_n) = \begin{cases} 1, & \text{if } \forall_{1\leq i\leq n}\, x_i = 1, \\ 0, & \text{if } \forall_{1\leq i\leq n}\, x_i = 0, \\ \frac{1}{2}, & \text{otherwise.} \end{cases} \tag{63}$$

We will limit our considerations only to the operators that are *reducible to interval*. In that case, logic-based uncertainty-aware similarity measure can be calculated in two steps:

1. compute Φ_i for each $u_i \in U$ using (55),
2. compute $[s](\mathbf{A}, \mathbf{B})$ as a finite sum of intervals using (56).

First step can be effectively implemented using two dimensional segment tree [3,16]. First, given the function $\Psi : [0, 1] \times [0, 1] \to [0, 1]$ we build the static two-dimensional segment tree that can answer Range Minimum and Maximum queries. The complexity of this preprocessing is $O(\beta^2)$ both in terms of time and memory. Then, because each Φ_i is a sum of k intervals, it can be calculated using k minimum and maximum queries ($O(\log^2 \beta)$ each). This in total requires $O(\beta^2 + nk \log^2 \beta)$ operations. An important observation is that after this step, we have $O(\beta^2)$ different possible values in Φ_i intervals.

This can be further simplified thanks to the (Ψ2) property. One may observe that

$$\left\{\Psi(a, b) \,:\, a \in [\underline{a}, \overline{a}], b \in [\underline{b}, \overline{b}]\right\} = \Big[\min(\Psi(\underline{a}, \overline{b}), \Psi(\overline{a}, \underline{b}),\ x\Big], \tag{64}$$

where

$$x = \begin{cases} \Psi(\underline{a}, \overline{b}), & \overline{b} < \underline{a}, \\ \Psi(\overline{a}, \underline{b}), & \overline{a} < \underline{b}, \\ \max\left\{\Psi(x, x) \,:\, \max(\underline{a}, \underline{b}) \leq x \leq \min(\overline{a}, \overline{b})\right\}, & \text{otherwise.} \end{cases} \tag{65}$$

This allows to use one-dimensional segment tree to calculate the required value.

At this point we can use naive approach to calculate $[s](\mathbf{A}, \mathbf{B})$ which will result in $O(\beta^2 + nk \log^2 \beta + k^n)$ complexity. Iterating through all possible combinations of intervals that make up all Ψ_i to calculate the $[s](\mathbf{A}, \mathbf{B})$ directly can be very time consuming, especially when $k \geq 2$ and n is large. Unfortunately, it turns out that optimizing this procedure is difficult in the general case. We can also look at the problem from the other side. Instead of trying to directly calculate the value of $[s](\mathbf{A}, \mathbf{B})$, we can try to answer the question whether given $a \in [0, 1]$ belongs to $[s](\mathbf{A}, \mathbf{B})$. From the adopted assumptions, we have only β^2 values to check. Besides, the optimization problem has been replaced by a decision problem.

Unfortunately, in the case of arithmetic mean being the aggregation operator, checking whether $a \in [s](\mathbf{A}, \mathbf{B})$ is equivalent to the subset sum problem, which is NP-complete. However, if we use the median, minimum, or maximum, then all assumptions are still met, and at the same time, a much faster calculation is possible. Besides, the median has several unique properties useful in many applications, while still being very similar to the arithmetic mean.

Algorithm 1 Procedure that computes $[s](\mathbf{A}, \mathbf{B})$ for median as aggregation operator.

Input: Φ_i for each $u_i \in U$ represented as a list of intervals, $|U| = n$ is odd
Output: $\left\{\mathcal{M}(\phi_1, \phi_2, \cdots, \phi_n) : \forall_{u_i \in U}\ \phi_i \in \Phi_i, \right\}$ as a list of intervals
1: $start \leftarrow$ NULL
2: $previous \leftarrow -1$
3: $intervals \leftarrow \emptyset$
4: **for** $a \in [0, 1]$ **do**
5: $\overline{L} \leftarrow \{i : 1 \leq i \leq n, \underline{\Phi}_i > a\}$
6: $\underline{L} \leftarrow \{i : 1 \leq i \leq n, \overline{\Phi}_i < a\}$
7: **if** $2 \cdot MAX(|\underline{L}|, |\overline{L}|) \leq n - 1$ **and** $a \in \Phi_i$ for some $i \notin \underline{L} \cup \overline{L}$ **then**
8: **if** $start$ = NULL **then**
9: $start \leftarrow a$
10: **end if**
11: **else**
12: **if** $start \neq$ NULL **then**
13: $intervals \leftarrow intervals \cup \left\{[start, previous]\right\}$
14: $start \leftarrow$ NULL
15: **end if**
16: **end if**
17: $previous \leftarrow a$
18: **end for**
19: **return** $intervals$

To show our approach, we will solve the problem for the median. Analogous methods allow calculating results for minimum and maximum. The naive approach to answer whether $a \in [s](\mathbf{A}, \mathbf{B})$ involves iteration through all $O(k^n)$ intervals. This can be done much faster thanks to the observation that

$$a \in [s](\mathbf{A}, \mathbf{B}) \text{ if and only if } a = \mathcal{M}(\phi_1, \phi_2, \cdots, \phi_n) \text{ for some } \phi_i \in \Phi_i. \tag{66}$$

Hence, if n is odd, a need to satisfy only two conditions:

1. there exist index i such that $a \in \Phi_i$ and
2. three disjoint sets of indexes $\underline{L} \cup L \cup \overline{L} = \{1, \cdots, n\} \setminus \{i\}$ exist, such that $\forall_{l \in \underline{L}}\ \overline{\Phi}_l < a$, $\forall_{l \in \overline{L}}\ \underline{\Phi}_l > a$ and

$$\left||\overline{L}| - |\underline{L}|\right| \leq |L| \tag{67}$$

If n is even, a need to satisfy:

1. there exists a pair of indexes i, j such that $a \in \left\{\frac{x+y}{2} : x \in \Phi_i, y \in \Phi_j\right\}$ and
2. three disjoint sets of indexes $\underline{L} \cup L \cup \overline{L} = \{1, \cdots, n\} \setminus \{i, j\}$ exist, such that $\forall_{l \in \underline{L}}\ \overline{\Phi}_l < a$, $\forall_{l \in \overline{L}}\ \underline{\Phi}_l > a$ and

$$\left||\overline{L}| - |\underline{L}|\right| \leq |L| \tag{68}$$

This observation allows to directly construct the Algorithm 1 that will check for all $a \in [0, 1]$ whether $a \in [s](\mathbf{A}, \mathbf{B})$ with complexity of $O(nk\beta^2)$. When n is even, we need to modify the condition in step 7 to check whether $MAX(|\underline{L}|, |\overline{L}|) \leq n - 2$ and $a \in \left\{\frac{x+y}{2} : x \in \Phi_i, y \in \Phi_j\right\}$ for some $i, j \notin \underline{L} \cup \overline{L}$ which, in total requires $O(n^2k^2\beta^2)$ operations.

The proposed algorithm can be implemented in such a way to avoid computing $\underline{L}$ and $\overline{L}$ from the definition in each step (we can update previous sets in constant time). Checking, whether $a \in \Phi_i$, can also be implemented using the Interval Tree data structure to require $O(\log nk)$ operations.

Very similar reasoning leads to the algorithm for the case of minimum (and maximum) aggregation operator:

$$a \in [s](\mathbf{A}, \mathbf{B}) \text{ if and only if } a = \min(\phi_1, \phi_2, \cdots, \phi_n) \text{ for some } \phi_i \in \Phi_i. \tag{69}$$

Hence, a needs to satisfy:

1. there exist index i such that $a \in \Phi_i$ and

2. $\forall_i\ a \leq \overline{\Phi}_i$,

Thanks to this, $a \in [s](\mathbf{A}, \mathbf{B})$ can be computed in $O(nk \log nk)$ operations, by calculating minimum of $\overline{\Phi}_i$ and finding an union of all nk intervals truncated by this minimum.

6. Conclusions

In the first part of the paper, we discussed our research's motivation, which is the measurement of similarity under epistemic uncertainty. Our approach gives a full picture of the similarity of incompletely known information. It allows reasoning about the amount of uncertainty of compared objects, thus informing about this comparison's quality. The properties postulated in previous works led to the set-valued extension computation problem. We showed that those two approaches: axiomatic proposed by the *uncertainty-aware* similarity measure and the constructive one with set-valued extension coincide.

The second part of the paper (Sections 4 and 5) was focused on the computational aspects of set-valued extensions of similarity measures. We managed to obtain some promising results that allow us to compute similarity effectively. The main focus was put on the fourth class of $\mathcal{F}(U)$ subsets – Typical Interval-Valued Hesitant Fuzzy Sets (TIVHFS). For all analyzed similarity measures, the bounding interval (min/max extension) can be computed in log-linear time.

Particularly interesting results were obtained for the family of logic-based (also called additive) similarity measures, which use the interpretation of the membership function as the degree of truth. Thanks to their unique properties, in the IoFS case, it is possible to compute them in linear time, regardless of the equality index and aggregation operator used. Nevertheless, calculations in the fourth case are still challenging, and further assumptions need to be made. The selection of the aggregation operator for logic-based similarity measure has a significant impact on the computing capabilities in the TIVHFS case. The use of arithmetic mean often leads to a subset sum problem, which is known to be NP-complete. Other aggregation operators were investigated to overcome this issue. Thanks to the use of a median, minimum, or maximum operator, we obtained polynomial complexity algorithms for exact value computation in the case of TIVHFS.

As for the areas of further research, we primarily indicate the work on the computational evaluation of the proposed methods in real-life decision-making problems. Work is currently underway on the use of uncertainty-aware similarity measures to support medical diagnostics [44], where epistemic data uncertainty is common. We also work on providing reference implementations of the proposed similarity measures that can be used in other projects. Research on the use of these measures in other areas of application will be highly desirable. One of the fundamental problems is to indicate the usefulness of individual equality values in various areas of application.

Another topic of further research is an attempt to provide the necessary conditions that Φ need to fulfil so that s_Ψ is a similarity measure. The second problem concerns the possibility of defining the uncertainty-aware similarity measure without using the set-valued extension or providing the general characteristics of all such measures. In this work, we focused on the computational utility of the proposed methods. Further theoretical analysis of those concepts can contribute to a better understanding of them and, as a result, further improve computational efficiency.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

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

References

[1] M. Baczyński, B. Jayaram, Fuzzy Implications, Studies in Fuzziness and Soft Computing, vol. 231, Springer, Heidelberg, 2008.
[2] G. Beliakov, A. Pradera, T. Calvo, et al., Aggregation Functions: A Guide for Practitioners, Studies in Fuzziness and Soft Computing, vol. 221, Springer-Verlag, Heidelberg, 2007.
[3] J.L. Bentley, Solutions to Klee's rectangle problems, 1977, pp. 282–300, unpublished manuscript.

[4] J.L. Bentley, T.A. Ottmann, Algorithms for reporting and counting geometric intersections, IEEE Trans. Comput. C-28 (9) (1979) 643–647.
[5] R.P. Brent, Algorithms for Minimization Without Derivatives, Dover Publications, 2013.
[6] H. Bustince, Indicator of inclusion grade for interval-valued fuzzy sets. Application to approximate reasoning based on interval-valued fuzzy sets, Int. J. Approx. Reason. 23 (3) (2000) 137–209.
[7] N. Chen, Z. Xu, M. Xia, Interval-valued hesitant preference relations and their applications to group decision making, Knowl.-Based Syst. 37 (2013) 528–540.
[8] S.-M. Chen, Measures of similarity between vague sets, Fuzzy Sets Syst. 74 (2) (1995) 217–223.
[9] I. Couso, H. Bustince, Three categories of set-valued generalizations from fuzzy sets to interval-valued and Atanassov intuitionistic fuzzy sets, IEEE Trans. Fuzzy Syst. 26 (5) (2017) 3112–3121.
[10] I. Couso, H. Bustince, From fuzzy sets to interval-valued and Atanassov intuitionistic fuzzy sets: a unified view of different axiomatic measures, IEEE Trans. Fuzzy Syst. 27 (2) (2019) 362–371.
[11] I. Couso, D. Dubois, Statistical reasoning with set-valued information: ontic vs. epistemic views, Int. J. Approx. Reason. 55 (7) (2014) 1502–1518.
[12] I. Couso, L. Garrido, L. Sánchez, Similarity and dissimilarity measures between fuzzy sets: a formal relational study, Inf. Sci. 229 (2013) 122–141.
[13] I. Couso, L. Sánchez, Additive similarity and dissimilarity measures, Fuzzy Sets Syst. 322 (2017) 35–53.
[14] I. Couso, L. Sánchez, A note on "Similarity and dissimilarity measures between fuzzy sets: a formal relational study" and "Additive similarity and dissimilarity measures", Fuzzy Sets Syst. 390 (2020) 183–187.
[15] V.V. Cross, T.A. Sudkamp, Similarity and Compatibility in Fuzzy Set Theory. Assessment and Applications, Physica-Verlag, Heidelberg, 2002.
[16] M. de Berg, O. Cheong, M. van Kreveld, M. Overmars, Computational Geometry: Algorithms and Applications, Chap. More Geometric Data Structures, Springer, Berlin, 2008, pp. 219–241.
[17] D. Dubois, H. Prade, Gradualness, uncertainty and bipolarity: making sense of fuzzy sets, Fuzzy Sets Syst. 192 (2012) 3–24.
[18] J. Fodor, M. Roubens, Fuzzy Preference Modelling and Multicriteria Decision Support, Theory and Decision Library, vol. 14, Springer Science & Business Media, 1994.
[19] 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.
[20] K. Hirota, W. Pedrycz, Matching fuzzy quantities, IEEE Trans. Syst. Man Cybern. 21 (6) (1991) 1580–1586.
[21] 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.
[22] A. Kandel, Fuzzy Mathematical Techniques with Applications, Addison-Wesley Publishing Co., Boston, 1986.
[23] N.N. Karnik, J.M. Mendel, Centroid of a type-2 fuzzy set, Inf. Sci. 132 (1) (2001) 195–220.
[24] E.P. Klement, R. Mesiar, E. Pap, Triangular Norms, Trends in Logic, vol. 8, Kluwer Academic Publishers, Dordrecht, 2000.
[25] Z. Liang, P. Shi, Similarity measures on intuitionistic fuzzy sets, Pattern Recognit. Lett. 24 (15) (2003) 2687–2693.
[26] X. Luo, C. Zhang, An axiom foundation for uncertain reasonings in rule-based expert systems: NT-algebra, Knowl. Inf. Syst. 1 (4) (1999) 415–433.
[27] J.M. Mendel, R.I. John, F. Liu, Interval type-2 fuzzy logic systems made simple, IEEE Trans. Fuzzy Syst. 14 (6) (2006) 808–821.
[28] H.B. Mitchell, On the Dengfeng–Chuntian similarity measure and its application to pattern recognition, Pattern Recognit. Lett. 24 (16) (2003) 3101–3104.
[29] O. Nempont, J. Atif, I. Bloch, A constraint propagation approach to structural model based image segmentation and recognition, Inf. Sci. 246 (2013) 1–27.
[30] H.T. Nguyen, V. Kreinovich, Computing degrees of subsethood and similarity for interval-valued fuzzy sets: fast algorithms, Tech. Rep. 94, Department of Computer Science, UTEP, 2008.
[31] A. Stachowiak, K. Dyczkowski, A similarity measure with uncertainty for incompletely known fuzzy sets, in: Proceedings of Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), Edmonton, Canada, 2013, pp. 390–394.
[32] 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, L. Doukovska, M. Hadjiski, V. Jotsov, J. Kacprzyk, N. Kasabov, S. Sotirov, E. Szmidt, S. Zadrożny (Eds.), Intelligent Systems'2014, in: Advances in Intelligent Systems and Computing, vol. 322, Springer, Cham, 2015, pp. 741–751.
[33] E. Szmidt, Distances and Similarities in Intuitionistic Fuzzy Sets, Springer, Switzerland, 2014.
[34] V. Torra, Hesitant fuzzy sets, Int. J. Intell. Syst. 25 (6) (2010) 529–539.
[35] 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.
[36] D. Wu, J.M. Mendel, Efficient algorithms for computing a class of subsethood and similarity measures for interval type-2 fuzzy sets, in: Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Barcelona, Spain, 2010, pp. 1–7.
[37] L. Xuecheng, Entropy, distance measure and similarity measure of fuzzy sets and their relations, Fuzzy Sets Syst. 52 (3) (1992) 305–318.
[38] L.A. Zadeh, Similarity relations and fuzzy orderings, Inf. Sci. 3 (2) (1971) 177–200.
[39] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning—I, Inf. Sci. 8 (3) (1975) 199–249.
[40] W. Zeng, H. Li, Relationship between similarity measure and entropy of interval valued fuzzy sets, Fuzzy Sets Syst. 157 (11) (2006) 1477–1484.
[41] H. Zhang, W. Zhang, C. Mei, Entropy of interval-valued fuzzy sets based on distance and its relationship with similarity measure, Knowl.-Based Syst. 22 (6) (2009) 449–454.
[42] 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, 2016 (in Polish).

[43] P. Żywica, Modelling medical uncertainties with use of fuzzy sets and their extensions, in: J. Medina, M. Ojeda-Aciego, J.L. Verdegay, I. Perfilieva, B. Bouchon-Meunier, R.R. Yager (Eds.), Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications. IMPU 2018, in: Communications in Computer and Information Science, vol. 855, Springer, Cham, 2018, pp. 369–380.

[44] P. Żywica, Application of uncertainty-aware similarity measure to classification in medical diagnosis, in: 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, United Kingdom, 2020, pp. 1–8.

[45] P. Żywica, A. Stachowiak, Uncertainty-aware similarity measures - properties and construction method, in: V. Novák, V. Mařík, M. Štěpnička, M. Navara, P. Hurtík (Eds.), 2019 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (EUSFLAT 2019), vol. 1, Atlantis Press, 2019, pp. 512–519.

[46] P. Żywica, A. Stachowiak, M. Wygralak, An algorithmic study of relative cardinalities for interval-valued fuzzy sets, Fuzzy Sets Syst. 294 (2015) 105–124.