A Similarity Measure Based on Bidirectional Subsethood for Intervals

Shaily Kabir, *Student Member, IEEE,* Christian Wagner, *Senior Member, IEEE,*
Timothy C. Havens, *Senior Member, IEEE,* and Derek T. Anderson, *Senior Member, IEEE*

Abstract—With a growing number of areas leveraging interval-valued data—including in the context of modelling human uncertainty (e.g., in cybersecurity), the capacity to accurately and systematically compare intervals for reasoning and computation is increasingly important. In practice, well established set-theoretic similarity measures such as the Jaccard and Sørensen-Dice measures are commonly used, while axiomatically a wide breadth of possible measures have been theoretically explored. This paper identifies, articulates, and addresses an inherent and so far not discussed limitation of popular measures—their tendency to be subject to *aliasing*—where they return the same similarity value for very different sets of intervals. The latter risks counter-intuitive results and poor automated reasoning in real-world applications dependent on systematically comparing interval-valued system variables or states. Given this, we introduce new axioms establishing desirable properties for robust similarity measures, followed by putting forward a novel set-theoretic similarity measure based on the concept of bidirectional subsethood which satisfies both traditional and new axioms. The proposed measure is designed to be sensitive to the variation in the size of intervals, thus avoiding *aliasing*. The paper provides a detailed theoretical exploration of the new proposed measure, and systematically demonstrates its behaviour using an extensive set of synthetic and real-world data. Specifically, the measure is shown to return robust outputs that follow intuition—essential for real world applications. For example, we show that it is bounded above and below by the Jaccard and Sørensen-Dice similarity measures (when the minimum *t-norm* is used). Finally, we show that a dissimilarity or distance measure, which satisfies the properties of a metric, can easily be derived from the proposed similarity measure.

Index Terms—similarity measure, distance measure, subsethood, interval-valued data

I. INTRODUCTION

SIMILARITY measures (SMs) are widely used in many areas including decision making, data aggregation, approximate reasoning, and machine learning. Various SMs have come into use to capture likeness among objects, though each of the measures has its own strengths and weaknesses. Similarity is commonly represented by a non-negative real number, often between 0, meaning objects are not similar

S. Kabir is with the School of Computer Science, University of Nottingham, Nottingham, UK (e-mail: shaily.kabir@nottingham.ac.uk).

C. Wagner is with the School of Computer Science, University of Nottingham, Nottingham, UK (e-mail: christian.wagner@nottingham.ac.uk).

T.C. Havens is with the Department of Electrical and Computer Engineering, and the Department of Computer Science, Michigan Technological University, Houghton, MI, USA (e-mail: thavens@mtu.edu).

D.T. Anderson is with the Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA (e-mail: andersondt@missouri.edu).
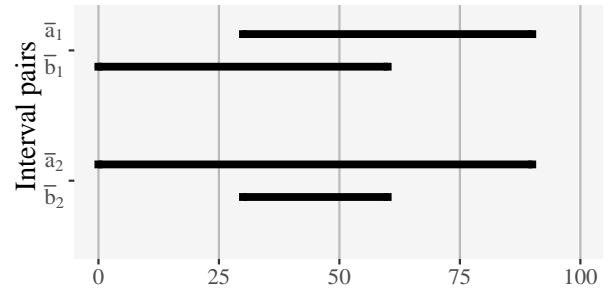
Fig. 1. Two different pairs of intervals showing *aliasing* with the Jaccard and Dice similarity measures

at all, and 1, meaning they are identical. SMs are generally symmetrical, though, for certain objects, similarity can be better modelled by unidirectional or asymmetric functions [1]. In addition, similarity measures can be transformed to capture distance using various functions [2], showing dissimilarity among objects (e.g., color images as in [3]).

Interval-valued data have recently gained much interest for the modelling of uncertainty and vagueness, particularly in the modelling of survey data [4], the representation of symbolic data [5], and the capture of natural language expressions [6], as they offer simple and efficient representation of uncertain, vague, and imprecise information. In such areas, intervals are often compared through various SMs. Among them, the Jaccard [7] and the Sørensen-Dice [8] SMs—henceforth referred to as Dice SM for convenience—are the most commonly used set-theoretic SMs in practice. These two measures provide symmetrical similarity, growing slowly from a minimum (0) to a maximum value (1) in response to an increasing degree of overlap between two closed intervals.

Nevertheless, both of these, and indeed most set-theoretic SMs frequently suffer from a so-far not discussed pitfall, best described as *aliasing*. *Aliasing* occurs where the same similarity value is generated for very different interval pairs. Figure 1 shows an example of such interval pairs for which the Jaccard and Dice SMs give the same similarity of 0.33 and 0.50 respectively. While SMs returning such identical results for different sets of intervals is at least counter-intuitive, at worst it leads to incorrect inference in real-world applications. The reason why these measures exhibit *aliasing* is because their sensitivity in respect to changes in the relative size of the intervals is limited, that is, they are largely driven by the size of intersection and union. However, it is reasonable to expect that similarity would vary *both* in respect to the overlap and to the mutual similarity-in-size of the intervals.

TABLE I
ACRONYMS AND NOTATION

SM	Similarity Measure
DM	Distance Measure
S_J	Jaccard SM
S_D	Dice SM
S_T	Tversky's ratio model [1]
S_h	Subsethood
S_{S_h}	Bidirectional subsethood based SM
D_{S_h}	Bidirectional subsethood based DM
X	Universe of discourse
a	Crisp set
$\bar{a}$	Interval $\{\bar{a} \subseteq \mathbb{R} : \bar{a} = [a^-, a^+],\ a^- < a^+\}$
A	Fuzzy set $\{(x, \mu_A(x)) | x \in X \text{ and } \mu_A(x) \in [0,1]\}$

We initially put forward the underpinnings of a new set-theoretic SM in [9] for pairs of (closed) intervals which considered their reciprocal overlapping ratios to capture similarity with a high degree of sensitivity.

This paper builds on our initial work [9], while using the term subsethood rather than overlapping ratio as they are equivalent in practice. Moreover, subsethood is commonly used and well understood in the literature. The major contributions of this paper are as follows–

• *Aliasing* is identified and articulated as a risk and potential shortcoming affecting many popular SMs.[1]

• Going beyond the common axioms for being a SM, five new axioms are introduced along with their justifications to expand the axiomatic definition towards robust SMs. Further, an axiomatic definition of subsethood for intervals is also provided (in Appendix).

• A new set-theoretic SM is proposed which is designed to be sensitive to the variation in the size of intervals, thus avoiding *aliasing*.

• The new SM is explored theoretically, showing its conformity with the expanded axiomatic definition, that is, both common and new axioms of a robust SM.

• A new distance measure (DM) is derived from the new SM for estimating distance or dissimilarity between intervals, and it is proved to be a metric.

• The utility of the new SM is demonstrated in the context of both synthetic and real world cases. Using both synthetic and real-world interval-valued data, we demonstrate its intuitive behaviour in line with popular SMs – specifically being bounded above and below by Jaccard and Dice – highlighting how the new SM follows expected and intuitive behaviour, while still addressing pitfalls of existing SMs (e.g., by avoiding *aliasing*).

The paper is structured as follows. Section II briefly reviews SMs, including the Jaccard and Dice, as well as DMs. Section III discusses an axiomatic definition towards robust SMs for (closed) intervals, followed by proposing a new set-theoretic SM which supports all axioms and then derives a new distance metric from the new SM. In Section IV, we explore

the behaviour of the new SM in respect to the Jaccard and Dice measures, using both synthetic and real world interval-valued datasets. Finally, Section V concludes the paper and briefly discusses future work. Table I presents a list of acronyms and notation used in this paper to assist the reader.

II. BACKGROUND

We review SMs generally, including their most common properties arising from their axiomatic definition, followed by a detailed review of the Jaccard and Dice measures as the most widely used SMs for intervals in the literature. Next, we briefly discuss DMs and the properties required for the latter being a metric, followed by a review of subsethood which is employed later in the paper to derive a new SM. We focus throughout on measures applied to intervals in this paper as the latter provide the foundation for the later extension to SMs for more complex data types, including fuzzy sets.

A. Similarity Measures

A SM $S(a, b) \rightarrow [0, 1]$ is a real-valued function that determines how two objects, a and b, are alike. Generally, the similarity between two objects is bounded by 0 and 1, where 0 means that both objects are completely different and 1 means that they are identical. The four common properties of a SM for sets a, b, and c are as follows [14]:

[A1] Boundedness: $0 \leq S(a, b) \leq 1$;
[A2] Symmetry: $S(a, b) = S(b, a)$;
[A3] Reflexivity: $S(a, b) = 1 \iff a = b$;
[A4] Transitivity: If $a \subseteq b \subseteq c$ then $S(a, b) \geq S(a, c)$.

The above properties are mirrored in (and in many cases arise from) the axiomatic definitions associated with a wide variety of established SMs. Going beyond these standard and generic properties, we now provide a more detailed review of both the Jaccard and Dice SMs as the most common SMs used in practice.

1) Jaccard Similarity Measure: The Jaccard SM [7] is one of the most widely used set-theoretic SMs, and it satisfies all of the above properties. Generally, the Jaccard similarity of two crisp sets a and b is defined as the ratio of the cardinality[2] of their intersection and the cardinality of their union,

$$S_J(a, b) = \frac{|a \cap b|}{|a \cup b|}. \tag{1}$$

Using the crisp set difference operation [15], Eq. (1) can be written as

$$S_J(a, b) = \frac{|a \cap b|}{|a \cap b| + |a \backslash b| + |b \backslash a|}, \tag{2}$$

where $a \backslash b$ is the set of items that are in a but not in b and $b \backslash a$ is the set of items that are in b but not in a. Equation (2) can also be derived from Tversky's parameterized ratio model of similarity [1]

$$S_T(a, b) = \frac{f(a \cap b)}{f(a \cap b) + \alpha f(a \backslash b) + \beta f(b \backslash a)},$$

---

[1]Beyond both Jaccard and Dice measures, other set-theoretic SMs such as Szymkiewicz-Simpson coefficient [10], Otsuka coefficient [11], Sokal-Sneath Coefficient [12], Simple-Matching coefficient [13] are also subject to *aliasing* for intervals.

[2]The cardinality of a set a is defined as the number of elements within a [15]. In this paper, we are interchangeably using the terms 'cardinality' and 'size' of a set. For intervals, we also use 'width'.

by setting the non-negative factors α and β to 1 and letting f be a cardinality function. Note that this alternative form of the Jaccard SM at Eq. (2) is relevant for showing its relationship with the Dice and other SMs, detailed in Section III. Beyond crisp sets, the Jaccard SM is used to estimate the similarity for intervals or sets of intervals such as employed for example in data fusion [16], [17] and that of fuzzy sets [18].

A closed interval $\overline{a}$ is a set of real numbers characterized by two endpoints a^- and a^+ with $a^- < a^+$.[3] The interval $\overline{a}$ is often represented as $[a^-, a^+]$ and its cardinality, size or width is $|\overline{a}| = |a^+ - a^-|$. For comparing the intervals $\overline{a}$ and $\overline{b}$, the Jaccard SM is expressed as

$$S_J\left(\overline{a}, \overline{b}\right) = \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cup \overline{b}\right|}, \tag{3}$$

where $|\overline{a}| \neq 0$ and $|\overline{b}| \neq 0$. $\left|\overline{a} \cap \overline{b}\right|$ is the size of the intersection between $\overline{a}$ and $\overline{b}$ and $\left|\overline{a} \cup \overline{b}\right|$ is the size of the interval segment(s) covering both $\overline{a}$ and $\overline{b}$. Hence, $S_J\left(\overline{a}, \overline{b}\right) = 1$ when $\overline{a}$ and $\overline{b}$ are completely overlapping and $S_J\left(\overline{a}, \overline{b}\right) = 0$ when they are not overlapping at all. Similar to Eq. (2), we can rewrite Eq. (3) as

$$S_J\left(\overline{a}, \overline{b}\right) = \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{a} \backslash \overline{b}\right| + \left|\overline{b} \backslash \overline{a}\right|}, \tag{4}$$

where $\left|\overline{a} \backslash \overline{b}\right|$ is the size of the interval segment of $\overline{a}$ that is not overlapping with $\overline{b}$ and $\left|\overline{b} \backslash \overline{a}\right|$ is the size of the interval segment of $\overline{b}$ that is not overlapping with $\overline{a}$.

A fuzzy set [21] is defined as a set where the set's elements have membership ranging between 0 and 1. Formally, a type-1 fuzzy set A on a discrete and finite universe of discourse X is written as [22]

$$A = \{(x, \mu_A(x)) \,|\, x \in X\}, \tag{5}$$

where $\mu_A(x) \in [0, 1]$ is the membership grade of the element x in A. For two type-1 fuzzy sets A and B on the discrete and finite universe X, the Jaccard SM can be written as [23]

$$S_J(A, B) = \frac{\sum\limits_{i=1}^{N} \min\left(\mu_A(x_i), \mu_B(x_i)\right)}{\sum\limits_{i=1}^{N} \max\left(\mu_A(x_i), \mu_B(x_i)\right)}, \tag{6}$$

where $\mu_A(x_i)$ and $\mu_B(x_i)$ are the membership grades of x_i in A and B respectively. Equation (6) yields a value of 1 when the fuzzy sets are identical and 0 when they are disjoint. It is noted that the Jaccard SM has been further extended for interval-valued fuzzy sets [24] [25] and type-2 fuzzy sets [26] [27]; though, this is not discussed further here.

2) Dice Similarity Measure: The Dice SM [8] is closely related to the Jaccard SM. To assess the similarity between two sets, it considers the ratio of the cardinality of their intersection and the average of their cardinality. Like the Jaccard similarity, it produces outputs in $[0, 1]$. Specifically, for two crisp sets a and b, the Dice similarity is expressed as

$$S_D(a, b) = \frac{|a \cap b|}{\frac{1}{2}(|a| + |b|)}, \tag{7}$$

where $|a|$ is the cardinality of the set a. We can rewrite Eq. (7) by applying the crisp set difference operation [15]

$$S_D(a, b) = \frac{|a \cap b|}{|a \cap b| + \frac{1}{2}(|a \backslash b| + |b \backslash a|)}. \tag{8}$$

Note that Eq. (8) can also be obtained from Tversky's ratio model [1] when both non-negative factors α and β are 0.5. The alternative expressions of Jaccard at Eq. (2) and Dice at Eq. (8) show clearly that the averaging operation in the denominator of Eq. (8) results in the Dice similarity always being equal to—when sets are identical—or larger than the Jaccard similarity. We expand on this in Section III.

In [16], [17], the Dice similarity is used along with the Jaccard similarity for intervals. By following Eq. (4), the Dice similarity for two intervals $\overline{a}$ and $\overline{b}$ can be expressed as

$$S_D\left(\overline{a}, \overline{b}\right) = \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \frac{1}{2}(\left|\overline{a} \backslash \overline{b}\right| + \left|\overline{b} \backslash \overline{a}\right|)}, \tag{9}$$

where $|\overline{a}| \neq 0$ and $|\overline{b}| \neq 0$. While less frequently used for fuzzy sets than Jaccard, the Dice SM is, for example, used in [28], [29] for trapezoidal fuzzy numbers in the context of solving multi-criteria decision-making problems.

B. Distance Measures

A DM $D(a, b) \rightarrow \mathbb{R}^+$ is a real-valued function that determines how far apart two objects a and b are. A DM is a metric when it satisfies the following properties for sets a, b, and c [30]:

[B1] Non-negativity: $D(a, b) \geq 0$;
[B2] Symmetry: $D(a, b) = D(b, a)$;
[B3] Identity: $D(a, b) = 0 \iff a = b$;
[B4] Triangle Inequality: $D(a, c) \leq D(a, b) + D(b, c)$.

The Jaccard DM is complementary to the Jaccard SM and is a distance metric [31]. It is simply obtained by $D_J(a, b) = 1 - S_J(a, b)$. The Dice DM is also obtained by subtracting the Dice SM from 1, i.e., $D_D(a, b) = 1 - S_D(a, b)$. However, it is not a distance metric, but is often referred to as a semimetric as it satisfies all of the preceding properties except the triangle inequality [31]. Note that a DM is often within the range $[0, 1]$—0 for identical objects and 1 for dissimilar objects—when it is derived from a similarity measure bounded by $[0, 1]$.

C. Subsethood

Subsethood between two crisp sets a and b is a relation that indicates the degree to which a is a subset of b [24]. It is defined as

$$S_h(a, b) = \frac{|a \cap b|}{|a|}, \tag{10}$$

where $|a \cap b|$ is the cardinality of the intersection of sets a and b, and $|a|$ is the cardinality of set a.[4] From Eq. (10), it is clear that subsethood is smaller when more elements of set a are not part of set b and is larger when more elements of

---

[3]Note that $\overline{a}$ is also known as a continuous [19] or convex [20] interval.

[4]Equation (10) can also be derived from Tversky's ratio model [1] by setting $\alpha = 1$ and $\beta = 0$.

set a are part of set b. In general, Eq. (10) is bounded on the interval $[0, 1]$, where 1 means that a is a subset of b ($a \subseteq b$) and 0 means a is not a subset of b ($a \not\subset b$)—they are disjoint.

In a similar manner, the degree of subsethood of two closed intervals $\overline{a}$ and $\overline{b}$ can be defined as

$$S_h\left(\overline{a}, \overline{b}\right) = \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a}\right|}, \tag{11}$$

where $\left|\overline{a} \cap \overline{b}\right|$ is the size of the intersection between $\overline{a}$ and $\overline{b}$ and $\left|\overline{a}\right| \neq 0$.[5]

A binary notion of subsethood for fuzzy sets is first defined in [21] where for two fuzzy sets A and B on the universe X, A is a subset of B ($A \subseteq B$) if and only if $\mu_A(x) \leq \mu_B, \forall x \in X$. This definition is inherently crisp—A is or is not a subset of B—which is incoherent in respect to the fuzzy set theory. Hence, many alternatives for fuzzy subsethood have been introduced. Using the set-theoretic approach, the degree to which A is a subset of B on a finite X is defined as [33]

$$S_h(A, B) = \begin{cases} 1, & A = \emptyset \\ \frac{\sum\limits_{i=1}^{n} \min(\mu_A(x_i), \mu_B(x_i))}{\sum\limits_{i=1}^{n} \mu_A(x_i)}, & A \neq \emptyset, \end{cases} \tag{12}$$

where $\sum_{i=1}^{n} \min\left(\mu_A\left(x_i\right), \mu_B\left(x_i\right)\right)$ is a measure of the cardinality of the intersection of membership functions of A and B, and $\sum_{i=1}^{n} \mu_A\left(x_i\right)$ is a measure of the cardinality of A. Contrarily, using the fuzzy implication operator $\mathscr{I}$, it is expressed as $S_h(A, B) = \inf_{x \in X} \mathscr{I}\left(\mu_A(x), \mu_B(x)\right)$ where $\mathscr{I} : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is such that $\mathscr{I}(0, 0) = \mathscr{I}(0, 1) = \mathscr{I}(1, 1) = 1$ and $\mathscr{I}(1, 0) = 0$ [34].

Further, fuzzy subsethood is characterized by different axiomatizations. Initially, four axioms are proposed to define it as a binary fuzzy relation in [35]. Subsequently, [36] offers another set of axioms for fuzzy subsethood. We note that these axioms (except the first two axioms) are equivalent to those in [35]. However, in [37] it is argued that some fuzzy subsethoods do not maintain all axioms and new ones are proposed. For fuzzy sets $A, B, C \in X$, these axioms are –

[C1] $S_h(A, B) = 1$ if and only if $A \subseteq B$;
[C2] If $P \leq A$, then $S_h(A, A^c) = 0$ if and only if $A = X$;
[C3] if $A \subseteq B \subseteq C$, then $S_h(C, A) \leq S_h(B, A)$ and if $A \subseteq B$, then $S_h(C, A) \leq S_h(C, B)$;

where P is the fuzzy set with $\mu_P(x) = \frac{1}{2}$ and A^c is the fuzzy set with $\mu_{A^c}(x) = 1 - \mu_A(x)$. Later, axiom [C3] is modified as "if $A \subseteq B \subseteq C$, then $S_h(C, A) \leq S_h(B, A)$ and $S_h(C, A) \leq S_h(C, B)$" in [38] for defining new fuzzy subsethoods and their application to cluster validity. Besides, axioms [C2] and [C3] are altered in [39] as "$S_h(A, A^c) = 0$ if and only if $A = X$" and "if $A \subseteq B$, then $S_h(A, C) \geq S_h(B, C)$ and $S_h(C, A) \leq S_h(C, B)$" respectively to introduce a new fuzzy DI-subsethood by aggregating implication functions.

In many cases, fuzzy subsethood axioms are extended for interval-valued, intuitionistic, and type-2 fuzzy sets. Specifically, [40], [41], and [42] present key axioms for interval-valued fuzzy set subsethood. Further, in [43] an interval-valued

fuzzy strong S-subsethood measure is defined by aggregating implication functions. The extension of fuzzy subsethood for intuitionistic [44] and type-2 [45] fuzzy sets are not discussed.

Remark 1. In this paper, we build on the existing literature on fuzzy subsethood, particularly its key axioms to develop main properties of subsethood for intervals, see Appendix.

D. Interaction of Subsethood and Similarity Measures for Fuzzy Sets

The relationship between subsethood and SM has been explored for fuzzy sets. Particularly, [46] and [47] establish interchangeability between fuzzy subsethood and fuzzy set SM based on their axiomatic definitions. [41] also studies similar relationship for interval-valued fuzzy sets. Further, a set of axioms are proposed in [48] to define the properties of a fuzzy set SM which is extensively used in earlier studies. Further, three different SMs for fuzzy values are proposed and compared in [49] where key properties are coincided with the axioms defined in [48]. Besides, [50] and [51] also consider the same set of axioms to design fuzzy set SMs for comparing images with a restricted equivalence function [52].

Remark 2. A variety of works have focused on definitions and extensions of subsethood across the fuzzy set literature, including for type-1 [33], interval-valued [40], and type-2 [45] fuzzy sets. In this paper, we focus on SMs for intervals rather than fuzzy sets as the latter provide the direct underpinnings for straightforward subsequent extension (via the alpha-cut decomposition representation) to fuzzy sets.

In the next section, we first introduce an expanded axiomatic definition for *robust* SMs, followed by proposing a new SM for closed intervals which fulfills said definition and provides robust comparison of intervals.

III. TOWARDS ROBUST SET-THEORETIC SIMILARITY MEASURES

As outlined in Section I, *aliasing* poses a so far not discussed challenge to the robust comparison of intervals. In order to address this and move towards robust SMs, we first put forward an axiomatic definition for robust SMs for (closed) intervals. Then, building on this, in Section III.B, we define a novel set-theoretic SM for intervals [9] based on their bidirectional subsethood, showing that it follows the axiomatic definition put forward. From this SM, we derive a new DM for intervals along with a proof showing the measure is a metric. It is worth mentioning that we also define the properties of the subsethood for intervals in respect to the wider literature, for conciseness this is provided in the Appendix.

A. An Expanded Axiomatic Definition for Robust Similarity Measures on Closed Intervals

A real-valued function $S : \overline{a} \times \overline{b} \rightarrow [0, 1]$ is defined as a similarity measure for (closed) intervals if it maintains the following axioms:

[P1] $0 \leq S(\overline{a}, \overline{b}) \leq 1$ (boundedness);
[P2] $S(\overline{a}, \overline{b}) = S(\overline{b}, \overline{a})$ (symmetry);

---

[5]It is noted that subsethood is also known as inclusion [32] and overlapping ratio [9] as it captures the overlapping ratio between intervals.

[P3] $S(\overline{a}, \overline{b}) = 1$ if and only if $\overline{a} = \overline{b}$ (reflexivity);

[P4] If $\overline{a} \subseteq \overline{b} \subseteq \overline{c}$, then $S(\overline{a}, \overline{b}) \geq S(\overline{a}, \overline{c})$ and $S(\overline{b}, \overline{c}) \geq S(\overline{b}, \overline{a})$ (transitivity);

[P5] $S(\overline{a}, \overline{b}) = 0$ if and only if $|\overline{a} \cap \overline{b}| = 0$ (disjointness);

[P6] For two interval pairs $\{\overline{a}_1, \overline{b}_1\}$ and $\{\overline{a}_2, \overline{b}_2\}$ where $|\overline{a}_1 \cap \overline{b}_1| > 0$, $|\overline{a}_2 \cap \overline{b}_2| > 0$, $|\overline{a}_1 \cup \overline{b}_1| = |\overline{a}_2 \cup \overline{b}_2|$, and $|\overline{a}_1 \cap \overline{b}_1| = |\overline{a}_2 \cap \overline{b}_2|$, $S(\overline{a}_1, \overline{b}_1) \neq S(\overline{a}_2, \overline{b}_2)$ if $|\overline{a}_1| \neq |\overline{a}_2|$ and $|\overline{a}_1| \neq |\overline{b}_2|$, and $|\overline{b}_1| \neq |\overline{a}_2|$ and $|\overline{b}_1| \neq |\overline{b}_2|$ (non-aliasing);

[P7] $S(\overline{a}, \overline{b}) < 1$ if $\overline{a} \subset \overline{b}$ (subsethood-limit);

[P8] For two interval pairs $\{\overline{a}_1, \overline{b}_1\}$ and $\{\overline{a}_2, \overline{b}_2\}$ with $|\overline{a}_1 \cap \overline{b}_1| > 0$ and $|\overline{a}_2 \cap \overline{b}_2| > 0$, $S(\overline{a}_1, \overline{b}_1) = S(\overline{a}_2, \overline{b}_2)$ if $\overline{a}_2 = n \times \overline{a}_1 = [n \times a_1^-, n \times a_1^+]$ and $\overline{b}_2 = n \times \overline{b}_1 = [n \times b_1^-, n \times b_1^+]$ where $n > 0$ is a scaling factor (scaling-invariance);

[P9] For two interval pairs $\{\overline{a}_1, \overline{b}_1\}$ and $\{\overline{a}_2, \overline{b}_2\}$ where $|\overline{a}_1| = |\overline{b}_1| = |\overline{a}_2| = |\overline{b}_2|$, $S(\overline{a}_1, \overline{b}_1) < S(\overline{a}_2, \overline{b}_2)$ if $|\overline{a}_1 \cap \overline{b}_1| < |\overline{a}_2 \cap \overline{b}_2|$ (increased-overlap).

Justification of the axioms: The axioms [P1]-[P4] are derived following the common axioms [A1]-[A4] stated in Section II.A. Axiom [P5] is put forward for disjoint intervals where their similarity should be 0 as they are not overlapping. The *aliasing* example discussed in Section I calls for the axiom [P6]. Axiom [P7] is presented for intervals when one interval is a proper subset of the other. Here, the similarity should be less than 1 as both intervals are not equal. Axiom [P8] is introduced for intervals where they are scaled up by a factor. In this case, the ratio of overlapping and non-overlapping segments stays the same while scaling; thus, the similarity between intervals should remain constant. Lastly, axiom [P9] suggests higher similarity for equal-sized intervals when their overlap increases.

B. Similarity Measure based on Bidirectional Subsethood

As noted in Section I, the motivation behind the new set-theoretic SM is to establish a SM which is sensitive to potentially (very) different cardinalities of the sets being compared, in particular when one is a subset of the other. Thus, the proposed SM takes into consideration the reciprocal subsethood (i.e., the overlapping ratio) of both sets/intervals being compared, in order to estimate their overall similarity.

Definition 1. The bidirectional subsethood based SM S_{S_h} for a pair of intervals, $\overline{a}$ and $\overline{b}$, is the *t-norm* of their reciprocal subsethoods $S_h(\overline{a}, \overline{b})$ and $S_h(\overline{b}, \overline{a})$, i.e.,

$$S_{S_h}(\overline{a}, \overline{b}) = \bigstar(S_h(\overline{a}, \overline{b}), S_h(\overline{b}, \overline{a})), \quad (13)$$

where $\bigstar$ is a *t-norm*.

In this paper, we use $\bigstar$ to refer to either minimum *t-norm* or product *t-norm*, while using '$\wedge$' and 'Π' to indicate the minimum and product *t-norm*s respectively. In future, we will explore other *t-norm*s beyond minimum and product.

Theorem 1. *Consider a minimum ($\wedge$) or product (Π) t-norm $\bigstar$ and the subsethood for intervals S_h. Then, $S_{S_h}(\overline{a}, \overline{b}) = \bigstar(S_h(\overline{a}, \overline{b}), S_h(\overline{b}, \overline{a}))$ is a SM for intervals $\overline{a}$ and $\overline{b}$ satisfying all axioms $[P1]$–$[P9]$.*

Proof: The proofs are given per axioms in Section III.C. ∎

Similar to Eqs. (4) and (9), we can rewrite Eq. (13) as

$$S_{S_h}(\overline{a}, \overline{b}) = \bigstar\left(\frac{|\overline{a} \cap \overline{b}|}{|\overline{a} \cap \overline{b}| + |\overline{a} \backslash \overline{b}|}, \frac{|\overline{a} \cap \overline{b}|}{|\overline{a} \cap \overline{b}| + |\overline{b} \backslash \overline{a}|}\right), \quad (14)$$

where $|\overline{a} \backslash \overline{b}|$ is the size of non-overlapping segment(s) of $\overline{a}$ as to $\overline{b}$ and vice versa for $|\overline{b} \backslash \overline{a}|$. Also, $|\overline{a}| \neq 0$ and $|\overline{b}| \neq 0$.

C. Properties of the Proposed Similarity Measure

We now explore the properties of the proposed bidirectional subsethood based SM $S_{S_h}(\overline{a}, \overline{b})$.[6]

Theorem 2 (Boundedness). $0 \leq S_{S_h}(\overline{a}, \overline{b}) \leq 1$.

Proof: All *t-norm*s are bounded by 0 and 1 [53]. This is also true for the subsethood S_h (see Appendix). This eventually means that $S_{S_h}(\overline{a}, \overline{b})$ always within $[0, 1]$, thus addressing the axiom [P1]. ∎

Theorem 3 (Symmetry). $S_{S_h}(\overline{a}, \overline{b}) = S_{S_h}(\overline{b}, \overline{a})$.

Proof: Any *t-norm* ($\bigstar$) is symmetric [53]. Therefore, $S_{S_h}(\overline{a}, \overline{b})$ is also symmetric. The S_{S_h} SM, thus, satisfies the axiom [P2]. ∎

Theorem 4 (Reflexivity). $S_{S_h}(\overline{a}, \overline{b}) = 1 \iff \overline{a} = \overline{b}$.

Proof: If $\overline{a} = \overline{b}$, then $S_h(\overline{a}, \overline{b}) = S_h(\overline{b}, \overline{a}) = 1$. From the boundary conditions of the *t-norm* ($\bigstar$) [53], $\bigstar(1, 1) = 1$, thus making $S_{S_h}(\overline{a}, \overline{b}) = 1$. Alternatively, $S_{S_h}(\overline{a}, \overline{b}) = 1$ means that both $S_h(\overline{a}, \overline{b})$ and $S_h(\overline{b}, \overline{a})$ are equal to 1. This only happens when $\overline{a}$ and $\overline{b}$ are identical intervals. Here, the S_{S_h} SM maintains the axiom [P3]. ∎

Theorem 5 (Transitivity). *If $\overline{a} \subseteq \overline{b} \subseteq \overline{c}$, then $S_{S_h}(\overline{a}, \overline{b}) \geq S_{S_h}(\overline{a}, \overline{c})$ and $S_{S_h}(\overline{b}, \overline{c}) \geq S_{S_h}(\overline{a}, \overline{c})$.*

Proof: if $\overline{a} \subseteq \overline{b} \subseteq \overline{c}$, then

$$S_{S_h}(\overline{a}, \overline{b}) = \bigstar\left(\frac{|\overline{a} \cap \overline{b}|}{|\overline{a}|}, \frac{|\overline{a} \cap \overline{b}|}{|\overline{b}|}\right) = \bigstar\left(\frac{|\overline{a}|}{|\overline{a}|}, \frac{|\overline{a}|}{|\overline{b}|}\right) = \frac{|\overline{a}|}{|\overline{b}|},$$

$$S_{S_h}(\overline{a}, \overline{c}) = \bigstar\left(\frac{|\overline{a} \cap \overline{c}|}{|\overline{a}|}, \frac{|\overline{a} \cap \overline{c}|}{|\overline{c}|}\right) = \bigstar\left(\frac{|\overline{a}|}{|\overline{a}|}, \frac{|\overline{a}|}{|\overline{c}|}\right) = \frac{|\overline{a}|}{|\overline{c}|}.$$

As $\overline{b} \subseteq \overline{c}$, it follows that $|\overline{b}| \leq |\overline{c}|$. Therefore, $\frac{|\overline{a}|}{|\overline{b}|} \geq \frac{|\overline{a}|}{|\overline{c}|}$ and hence, $S_{S_h}(\overline{a}, \overline{b}) \geq S_{S_h}(\overline{a}, \overline{c})$. Again,

$$S_{S_h}(\overline{b}, \overline{c}) = \bigstar\left(\frac{|\overline{b} \cap \overline{c}|}{|\overline{b}|}, \frac{|\overline{b} \cap \overline{c}|}{|\overline{c}|}\right) = \bigstar\left(\frac{|\overline{b}|}{|\overline{b}|}, \frac{|\overline{b}|}{|\overline{c}|}\right) = \frac{|\overline{b}|}{|\overline{c}|}.$$

As $\overline{a} \subseteq \overline{b}$, it follows that $|\overline{b}| \geq |\overline{a}|$. Hence, $\frac{|\overline{b}|}{|\overline{c}|} \geq \frac{|\overline{a}|}{|\overline{c}|}$, implying that $S_{S_h}(\overline{b}, \overline{c}) \geq S_{S_h}(\overline{a}, \overline{c})$. In this case, the new SM satisfies the axiom [P4]. ∎

Theorem 6 (Disjointness). $S_{S_h}(\overline{a}, \overline{b}) = 0$, *if and only if $\overline{a}$ and $\overline{b}$ are disjoint.*

---

[6]Note that all properties discussed in this paper are maintained by the proposed S_{S_h} SM with both minimum ($\wedge$) and product (Π) *t-norm*s, unless otherwise specified.

Proof: If $\bar{a}$ and $\bar{b}$ are disjoint, then $|\bar{a} \cap \bar{b}| = 0$. Hence, $S_h\left(\bar{a}, \bar{b}\right) = 0$ and $S_h\left(\bar{b}, \bar{a}\right) = 0$, implying that $S_{S_h}\left(\bar{a}, \bar{b}\right) = \star(0,0) = 0$. Again, $S_{S_h}\left(\bar{a}, \bar{b}\right) = 0$ only when both $S_h\left(\bar{a}, \bar{b}\right)$ and $S_h\left(\bar{b}, \bar{a}\right)$ are equal to 0. This only happens when $\bar{a}$ and $\bar{b}$ are not overlapping. Here, the proof of this theorem shows that the new SM follows the axiom [P5]. ∎

Theorem 7 (Non-aliasing). *For two interval pairs $\{\bar{a}_1, \bar{b}_1\}$ and $\{\bar{a}_2, \bar{b}_2\}$ where $|\bar{a}_1 \cap \bar{b}_1| > 0$, $|\bar{a}_2 \cap \bar{b}_2| > 0$, $|\bar{a}_1 \cup \bar{b}_1| = |\bar{a}_2 \cup \bar{b}_2|$, and $|\bar{a}_1 \cap \bar{b}_1| = |\bar{a}_2 \cap \bar{b}_2|$, $S_{S_h}\left(\bar{a}_1, \bar{b}_1\right) \neq S_{S_h}\left(\bar{a}_2, \bar{b}_2\right)$ if $|\bar{a}_1| \neq |\bar{a}_2|$ and $|\bar{a}_1| \neq |\bar{b}_2|$, and $|\bar{b}_1| \neq |\bar{a}_2|$ and $|\bar{b}_1| \neq |\bar{b}_2|$.*

Proof: For the pair $\{\bar{a}_1, \bar{b}_1\}$, $S_h\left(\bar{a}_1, \bar{b}_1\right) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}$ and $S_h\left(\bar{b}_1, \bar{a}_1\right) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}$. Thus,

$$S_{S_h}\left(\bar{a}_1, \bar{b}_1\right) = \star\left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}\right).$$

Again, for the pair $\{\bar{a}_2, \bar{b}_2\}$, $S_h\left(\bar{a}_2, \bar{b}_2\right) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}$ and $S_h\left(\bar{b}_2, \bar{a}_2\right) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|}$. Thus,

$$S_{S_h}\left(\bar{a}_2, \bar{b}_2\right) = \star\left(\frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}, \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|}\right).$$

As $|\bar{a}_1 \cap \bar{b}_1| = |\bar{a}_2 \cap \bar{b}_2|$, therefore,

$$S_{S_h}\left(\bar{a}_2, \bar{b}_2\right) \equiv \star\left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_2|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_2|}\right).$$

Given that $|\bar{a}_1| \neq |\bar{a}_2|$ and $|\bar{a}_1| \neq |\bar{b}_2|$, it implies that

$$\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|} \neq \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_2|} \text{ and } \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|} \neq \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_2|}.$$

Similarly, $|\bar{b}_1| \neq |\bar{a}_2|$ and $|\bar{b}_1| \neq |\bar{b}_2|$ meaning that

$$\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} \neq \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_2|} \text{ and } \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} \neq \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_2|}.$$

Therefore,

$$\star\left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}\right) \neq \star\left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_2|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_2|}\right).$$

Hence, $S_{S_h}\left(\bar{a}_1, \bar{b}_1\right) \neq S_{S_h}\left(\bar{a}_2, \bar{b}_2\right)$, thus satisfying the axiom [P6]. ∎

Theorem 8 (Subsethood-limit). $S_{S_h}\left(\bar{a}, \bar{b}\right) < 1$ if $\bar{a} \subset \bar{b}$.

Proof: If $\bar{a} \subset \bar{b}$, then $|\bar{a}| < |\bar{b}|$ and $|\bar{a} \cap \bar{b}| = |\bar{a}|$. Thus, $S_h\left(\bar{a}, \bar{b}\right) = \frac{|\bar{a} \cap \bar{b}|}{|\bar{a}|} = \frac{|\bar{a}|}{|\bar{a}|} = 1$ and $S_h\left(\bar{b}, \bar{a}\right) = \frac{|\bar{a} \cap \bar{b}|}{|\bar{b}|} = \frac{|\bar{a}|}{|\bar{b}|}$. It means that $S_{S_h}\left(\bar{a}, \bar{b}\right) = \star\left(1, \frac{|\bar{a}|}{|\bar{b}|}\right) = \frac{|\bar{a}|}{|\bar{b}|}$. As $|\bar{a}| < |\bar{b}|$, then $\frac{|\bar{a}|}{|\bar{b}|} < 1$. Hence, $S_{S_h}\left(\bar{a}, \bar{b}\right) < 1$. Here, the new SM, thus, addresses the axiom [P7]. ∎

Theorem 9 (Scaling-invariance). *For two interval pairs $\{\bar{a}_1, \bar{b}_1\}$ and $\{\bar{a}_2, \bar{b}_2\}$ with $|\bar{a}_1 \cap \bar{b}_1| > 0$ and $|\bar{a}_2 \cap \bar{b}_2| > 0$,* $S_{S_h}\left(\bar{a}_1, \bar{b}_1\right) = S_{S_h}\left(\bar{a}_2, \bar{b}_2\right)$ *if $\bar{a}_2 = n \times \bar{a}_1 = [n \times a_1^-, n \times a_1^+]$ and $\bar{b}_2 = n \times \bar{b}_1 = [n \times b_1^-, n \times b_1^+]$ where $n > 0$ is a scaling factor.*

Proof: For the pair $\{\bar{a}_1, \bar{b}_1\}$, $S_h\left(\bar{a}_1, \bar{b}_1\right) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}$ and $S_h\left(\bar{b}_1, \bar{a}_1\right) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}$. Thus,

$$S_{S_h}\left(\bar{a}_1, \bar{b}_1\right) = \star\left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}\right).$$

Again, for the pair $\{\bar{a}_2, \bar{b}_2\}$, $S_h\left(\bar{a}_2, \bar{b}_2\right) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}$ and $S_h\left(\bar{b}_2, \bar{a}_2\right) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|}$. Thus,

$$S_{S_h}\left(\bar{a}_2, \bar{b}_2\right) = \star\left(\frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}, \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|}\right).$$

Given that $\bar{a}_2 = n \times \bar{a}_1$ and $\bar{b}_2 = n \times \bar{b}_1$. It implies that $|\bar{a}_2 \cap \bar{b}_2| = |(n \times \bar{a}_1) \cap (n \times \bar{b}_1)| = n \times |\bar{a}_1 \cap \bar{b}_1|$. Therefore,

$$S_{S_h}\left(\bar{a}_2, \bar{b}_2\right) = \star\left(\frac{n \times |\bar{a}_1 \cap \bar{b}_1|}{n \times |\bar{a}_1|}, \frac{n \times |\bar{a}_1 \cap \bar{b}_1|}{n \times |\bar{b}_1|}\right)$$
$$= \star\left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}\right) = S_{S_h}\left(\bar{a}_1, \bar{b}_1\right).$$

Hence, $S_{S_h}\left(\bar{a}_1, \bar{b}_1\right) = S_{S_h}\left(\bar{a}_2, \bar{b}_2\right)$. Here, the new SM follows the axiom [P8]. ∎

Theorem 10 (Increased-overlap). *For two interval pairs $\{\bar{a}_1, \bar{b}_1\}$ and $\{\bar{a}_2, \bar{b}_2\}$ where $|\bar{a}_1| = |\bar{b}_1| = |\bar{a}_2| = |\bar{b}_2|$,* $S_{S_h}\left(\bar{a}_1, \bar{b}_1\right) < S_{S_h}\left(\bar{a}_2, \bar{b}_2\right)$ *if $|\bar{a}_1 \cap \bar{b}_1| < |\bar{a}_2 \cap \bar{b}_2|$.*

Proof: For the pair $\{\bar{a}_1, \bar{b}_1\}$, $S_h\left(\bar{a}_1, \bar{b}_1\right) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}$ and $S_h\left(\bar{b}_1, \bar{a}_1\right) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}$. Thus,

$$S_{S_h}\left(\bar{a}_1, \bar{b}_1\right) = \star\left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}\right).$$

For the pair $\{\bar{a}_2, \bar{b}_2\}$, $S_h\left(\bar{a}_2, \bar{b}_2\right) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}$ and $S_h\left(\bar{b}_2, \bar{a}_2\right) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|}$. Thus,

$$S_{S_h}\left(\bar{a}_2, \bar{b}_2\right) = \star\left(\frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}, \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|}\right).$$

As $|\bar{a}_1| = |\bar{b}_1| = |\bar{a}_2| = |\bar{b}_2|$, therefore,

$$S_{S_h}\left(\bar{a}_2, \bar{b}_2\right) \equiv \star\left(\frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_1|}, \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_1|}\right).$$

Given that $|\bar{a}_1 \cap \bar{b}_1| < |\bar{a}_2 \cap \bar{b}_2|$, it implies that $\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|} < \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_1|}$ and $\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} < \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_1|}$. Therefore,

$$\star\left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}\right) < \star\left(\frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_1|}, \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_1|}\right).$$

Hence, $S_{S_h}\left(\overline{a}_1, \overline{b}_1\right) < S_{S_h}\left(\overline{a}_2, \overline{b}_2\right)$. Here, the S_{S_h} SM addresses the axiom [P9]. ∎

Theorem 11. $S_{S_h}\left(\overline{a}, \overline{b}\right)$ *is bounded by the Jaccard and Dice SMs when* ★ *is the minimum* (∧) *t-norm. That is,* $S_J\left(\overline{a}, \overline{b}\right) \leq S_{S_h \wedge}\left(\overline{a}, \overline{b}\right) \leq S_D\left(\overline{a}, \overline{b}\right)$.

Proof: For the interval pair $\{\overline{a}, \overline{b}\}$, consider the formulations of the SMs at Eqs. (4), (9), and (14).

To prove this theorem, we consider four cases: 1) $\overline{a} = \overline{b}$, 2) $\overline{a} \cap \overline{b} = \emptyset$, 3) $\overline{a} \subset \overline{b}$, and 4) $\overline{a} \cap \overline{b} \neq \emptyset$, and $\overline{a} \not\subset \overline{b}$ and $\overline{b} \not\subset \overline{a}$.

Case 1: If $\overline{a} = \overline{b}$, then all three measures yield a similarity of 1. That is, $S_J\left(\overline{a}, \overline{b}\right) = S_D\left(\overline{a}, \overline{b}\right) = S_{S_h \wedge}\left(\overline{a}, \overline{b}\right) = 1$.

Case 2: If $\overline{a} \cap \overline{b} = \emptyset$, all three measures result in a similarity of 0. Thus, $S_J\left(\overline{a}, \overline{b}\right) = S_D\left(\overline{a}, \overline{b}\right) = S_{S_h \wedge\left(\overline{a}, \overline{b}\right)} = 0$.

Case 3: If $\overline{a} \subset \overline{b}$ (subset), then $\left|\overline{a} \cap \overline{b}\right| = \left|\overline{a}\right|$. With respect to $\overline{b}$, there is no non-overlap segment of $\overline{a}$; hence, $\left|\overline{a}\backslash\overline{b}\right| = 0$. Inversely, there is a non-overlap segment of $\overline{b}$ as to $\overline{a}$; thus, $\left|\overline{b}\backslash\overline{a}\right| \neq 0$. In this case, the three SMs can be simplified to

$$S_J\left(\overline{a}, \overline{b}\right) = \frac{\left|\overline{a}\right|}{\left|\overline{a}\right| + \left|\overline{b}\backslash\overline{a}\right|},$$

$$S_D\left(\overline{a}, \overline{b}\right) = \frac{\left|\overline{a}\right|}{\left|\overline{a}\right| + \frac{1}{2}\left|\overline{b}\backslash\overline{a}\right|},$$

$$S_{S_h \wedge}\left(\overline{a}, \overline{b}\right) = \wedge\left(\frac{\left|\overline{a}\right|}{\left|\overline{a}\right|}, \frac{\left|\overline{a}\right|}{\left|\overline{a}\right| + \left|\overline{b}\backslash\overline{a}\right|}\right) = \frac{\left|\overline{a}\right|}{\left|\overline{a}\right| + \left|\overline{b}\backslash\overline{a}\right|},$$

which implies that

$$S_J\left(\overline{a}, \overline{b}\right) = S_{S_h \wedge}\left(\overline{a}, \overline{b}\right) < S_D\left(\overline{a}, \overline{b}\right).$$

Case 4: If $\overline{a} \cap \overline{b} \neq \emptyset$, and $\overline{a} \not\subset \overline{b}$ and $\overline{b} \not\subset \overline{a}$ ($\overline{a}$ and $\overline{b}$ are partially overlapping), then assume the case $\left|\overline{a}\right| \leq \left|\overline{b}\right|$. It implies that $\left|\overline{a}\right| - \left|\overline{a} \cap \overline{b}\right| \leq \left|\overline{b}\right| - \left|\overline{a} \cap \overline{b}\right| \implies \left|\overline{a}\backslash\overline{b}\right| \leq \left|\overline{b}\backslash\overline{a}\right|$. The three SMs are

$$S_J\left(\overline{a}, \overline{b}\right) = \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{a}\backslash\overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|},$$

$$S_D\left(\overline{a}, \overline{b}\right) = \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \frac{1}{2}\left(\left|\overline{a}\backslash\overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|\right)},$$

$$S_{S_h \wedge(\overline{a}, \overline{b})} = \wedge\left(\frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{a}\backslash\overline{b}\right|}, \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|}\right)$$

$$= \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|}, \text{ as } \left|\overline{a}\backslash\overline{b}\right| \leq \left|\overline{b}\backslash\overline{a}\right|.$$

It is true that

$$\frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{a}\backslash\overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|} < \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \frac{1}{2}\left(\left|\overline{a}\backslash\overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|\right)},$$

thus $S_J\left(\overline{a}, \overline{b}\right) < S_D\left(\overline{a}, \overline{b}\right)$. Again, it is clear that

$$\frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{a}\backslash\overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|} < \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|},$$

implying that $S_J\left(\overline{a}, \overline{b}\right) < S_{S_h \wedge}\left(\overline{a}, \overline{b}\right)$. Also,

$$\frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|} = \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \frac{1}{2}\left|\overline{b}\backslash\overline{a}\right| + \frac{1}{2}\left|\overline{b}\backslash\overline{a}\right|}$$

$$\leq \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \frac{1}{2}\left|\overline{a}\backslash\overline{b}\right| + \frac{1}{2}\left|\overline{b}\backslash\overline{a}\right|}, \text{ as } \left|\overline{a}\backslash\overline{b}\right| \leq \left|\overline{b}\backslash\overline{a}\right|,$$

indicating that $S_{S_h \wedge}\left(\overline{a}, \overline{b}\right) \leq S_D\left(\overline{a}, \overline{b}\right)$. Hence, $S_J\left(\overline{a}, \overline{b}\right) < S_{S_h \wedge}\left(\overline{a}, \overline{b}\right) \leq S_D\left(\overline{a}, \overline{b}\right)$. Note that for the case $\left|\overline{b}\right| \leq \left|\overline{a}\right|$, the same procedure can be used to prove the above relation. ∎

D. Distance Measure based on Bidirectional Subsethood

A new DM $D_{S_h}(\overline{a}, \overline{b})$ can easily be derived from the S_{S_h} measure at Eq. (13) by taking its complement, capturing the dissimilarity between both intervals:

$$D_{S_h}\left(\overline{a}, \overline{b}\right) = 1 - S_{S_h}\left(\overline{a}, \overline{b}\right). \tag{15}$$

Alternatively, Eq. (15) can be written as

$$D_{S_h}(\overline{a}, \overline{b}) = 1 - \bigstar\left(\frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{a}\backslash\overline{b}\right|}, \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|}\right)$$

$$= \bigstar\left(1 - \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{a}\backslash\overline{b}\right|}, 1 - \frac{\left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|}\right)$$

$$= \bigstar\left(\frac{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{a}\backslash\overline{b}\right| - \left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{a}\backslash\overline{b}\right|}, \frac{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right| - \left|\overline{a} \cap \overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|}\right)$$

$$= \bigstar\left(\frac{\left|\overline{a}\backslash\overline{b}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{a}\backslash\overline{b}\right|}, \frac{\left|\overline{b}\backslash\overline{a}\right|}{\left|\overline{a} \cap \overline{b}\right| + \left|\overline{b}\backslash\overline{a}\right|}\right).$$

Note that this alternative form of the proposed $D_{S_h}\left(\overline{a}, \overline{b}\right)$ measure can now directly be used in pattern recognition problems with sets/intervals such as classification and clustering.

We next discuss the essential properties of the $D_{S_h}\left(\overline{a}, \overline{b}\right)$ measure in terms of it being a metric.

Theorem 12. $D_{S_h}\left(\overline{a}, \overline{b}\right) = 1 - S_{S_h}\left(\overline{a}, \overline{b}\right)$ *is a metric.*

Proof: To prove that the DM, $D_{S_h}\left(\overline{a}, \overline{b}\right)$ is a metric, we need to show that it satisfies the following properties for the intervals $\overline{a}$, $\overline{b}$, and $\overline{c}$ where ★ is the minimum (∧) or product (Π) *t-norm*:

(a) $D_{S_h}\left(\overline{a}, \overline{b}\right) \geq 0$ (non-negativity),
(b) $D_{S_h}\left(\overline{a}, \overline{b}\right) = D_{S_h}\left(\overline{b}, \overline{a}\right)$ (symmetry),
(c) $D_{S_h}\left(\overline{a}, \overline{b}\right) = 0 \iff \overline{a} = \overline{b}$ (identity),
(d) $D_{S_h}\left(\overline{a}, \overline{c}\right) \leq D_{S_h}\left(\overline{a}, \overline{b}\right) + D_{S_h}\left(\overline{b}, \overline{c}\right)$ (triangle inequality).

Below we provide proofs for all the properties (a–d) above.

(a) By Theorem 3, $S_{S_h}\left(\overline{a}, \overline{b}\right)$ is always bounded by $[0, 1]$. As $D_{S_h}\left(\overline{a}, \overline{b}\right) = 1 - S_{S_h}\left(\overline{a}, \overline{b}\right)$, it implies that $0 \leq D_{S_h}\left(\overline{a}, \overline{b}\right) \leq 1$, thus meeting the non-negativity property.

(b) By Theorem 4, $S_{S_h}\left(\overline{a}, \overline{b}\right)$ is always symmetric; thus, $D_{S_h}\left(\overline{a}, \overline{b}\right)$ is also symmetric.

(c) Theorem 5 shows that if $\overline{a} = \overline{b}$, then $S_{S_h}\left(\overline{a}, \overline{b}\right)$ is always 1. Hence, $1 - S_{S_h}\left(\overline{a}, \overline{b}\right)$ makes $D_{S_h}\left(\overline{a}, \overline{b}\right) = 0$. Alternatively, If $\overline{a} \neq \overline{b}$, then $S_{S_h}\left(\overline{a}, \overline{b}\right) \neq 1$, meaning $D_{S_h}\left(\overline{a}, \overline{b}\right) \neq 0$.

(d) For the intervals $\bar{a}$, $\bar{b}$, and $\bar{c}$, we consider $\bar{a} \subseteq \bar{b} \subseteq \bar{c}$. It implies that $|\bar{a}| \le |\bar{b}| \le |\bar{c}|$. To prove this theorem, we apply the formulation of the S_{S_h} measure at Eq. (14).

Case 1: When all three intervals are equal $\left(\bar{a} = \bar{b} = \bar{c}\right)$, $S_{S_h}\left(\bar{a},\bar{b}\right) = S_{S_h}\left(\bar{b},\bar{c}\right) = S_{S_h}\left(\bar{a},\bar{c}\right) = 1$. Accordingly, $D_{S_h}\left(\bar{a},\bar{b}\right) = D_{S_h}\left(\bar{b},\bar{c}\right) = D_{S_h}\left(\bar{a},\bar{c}\right) = 0$, thus satisfying the triangle inequality.

Case 2: When $\bar{a} \subset \bar{b} \subset \bar{c}$, it implies that $|\bar{a}| < |\bar{b}| < |\bar{c}|$. Now, we can calculate the similarity and distance for each pair of intervals by applying Eq. (14):

(i) For $\bar{a} \subset \bar{b}$,

$$S_{S_h}\left(\bar{a},\bar{b}\right) = \bigstar \left(\frac{|\bar{a} \cap \bar{b}|}{|\bar{a}|}, \frac{|\bar{a} \cap \bar{b}|}{|\bar{b}|} \right)$$

$$= \bigstar \left(\frac{|\bar{a}|}{|\bar{a}|}, \frac{|\bar{a}|}{|\bar{b}|} \right) = \frac{|\bar{a}|}{|\bar{b}|}$$

Thus, $D_{S_h}\left(\bar{a},\bar{b}\right) = 1 - S_{S_h}\left(\bar{a},\bar{b}\right) = 1 - \frac{|\bar{a}|}{|\bar{b}|}$.

(ii) For $\bar{b} \subset \bar{c}$,

$$S_{S_h}\left(\bar{b},\bar{c}\right) = \frac{|\bar{b}|}{|\bar{c}|}$$

Thus, $D_{S_h}\left(\bar{b},\bar{c}\right) = 1 - S_{S_h}\left(\bar{b},\bar{c}\right) = 1 - \frac{|\bar{b}|}{|\bar{c}|}$.

(iii) For $\bar{a} \subset \bar{c}$ (by the transitive property of subsets [15]),

$$S_{S_h}\left(\bar{a},\bar{c}\right) = \frac{|\bar{a}|}{|\bar{c}|}$$

Thus, $D_{S_h}\left(\bar{a},\bar{c}\right) = 1 - S_{S_h}\left(\bar{a},\bar{c}\right) = 1 - \frac{|\bar{a}|}{|\bar{c}|}$.

Now, we have to show that

$$D_{S_h}\left(\bar{a},\bar{c}\right) < D_{S_h}\left(\bar{a},\bar{b}\right) + D_{S_h}\left(\bar{b},\bar{c}\right).$$

By placing the distance of each pair of intervals in the above equation, we get the following:

$$1 - \frac{|\bar{a}|}{|\bar{c}|} < 1 - \frac{|\bar{a}|}{|\bar{b}|} + 1 - \frac{|\bar{b}|}{|\bar{c}|}$$

$$\iff 1 - \frac{|\bar{a}|}{|\bar{c}|} - 1 + \frac{|\bar{b}|}{|\bar{c}|} < 1 - \frac{|\bar{a}|}{|\bar{b}|}$$

$$\iff -\frac{|\bar{a}|}{|\bar{c}|} + \frac{|\bar{b}|}{|\bar{c}|} < 1 - \frac{|\bar{a}|}{|\bar{b}|}$$

$$\iff \frac{|\bar{b}| - |\bar{a}|}{|\bar{c}|} < \frac{|\bar{b}| - |\bar{a}|}{|\bar{b}|}$$

$$\iff \frac{1}{|\bar{c}|} < \frac{1}{|\bar{b}|}.$$

It is true as $|\bar{b}| < |\bar{c}|$, thereby satisfying the triangle inequality. Thus, $D_{S_h}\left(\bar{a},\bar{c}\right) \le D_{S_h}\left(\bar{a},\bar{b}\right) + D_{S_h}\left(\bar{b},\bar{c}\right)$. ∎

Example satisfying triangle inequality: Consider three intervals $\bar{a}$, $\bar{b}$, and $\bar{c}$ in Fig. 2 where $\bar{b}, \bar{c} \subset \bar{a}$ and $\bar{b} \cap \bar{c} = \emptyset$. Table II
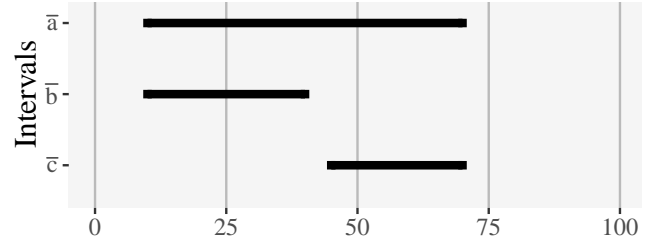

Fig. 2. Example that shows to satisfy the triangle inequality by the D_{S_h} measure

TABLE II
SIMILARITY AND DISTANCE BETWEEN THE INTERVALS OF FIG. 2

Interval Pair	$S_{S_h \wedge}$	$S_{S_h \Pi}$	$D_{S_h \wedge}$	$D_{S_h \Pi}$
$\{\bar{a},\bar{b}\}$	0.50	0.50	0.50	0.50
$\{\bar{a},\bar{c}\}$	0.42	0.42	0.58	0.58
$\{\bar{b},\bar{c}\}$	0.0	0.0	1.0	1.0

presents the similarity and distance between each pair of the three intervals using the S_{S_h} and D_{S_h} measures respectively with the minimum ($\wedge$) and product (Π) *t-norm*s. Note that the same similarity and distance results are received using both *t-norm*s. From the distance results, it is observed that for the interval pair $\{\bar{a},\bar{c}\}$, their distance—which is 0.58—is less than the summed distances of the pairs $\{\bar{a},\bar{b}\}$ and $\{\bar{b},\bar{c}\}$—which is 1.50—implying $D_{S_h}\left(\bar{a},\bar{c}\right) < D_{S_h}\left(\bar{a},\bar{b}\right) + D_{S_h}\left(\bar{b},\bar{c}\right)$. This relation is also maintained for the interval pairs $\{\bar{a},\bar{b}\}$ and $\{\bar{b},\bar{c}\}$, thereby demonstrating the D_{S_h} measure meeting the triangle inequality.

Theorem 13. $D_{S_h}\left(\bar{a},\bar{b}\right)$ *follows the property of transitivity. That is, $D_{S_h}\left(\bar{a},\bar{b}\right) \le D_{S_h}\left(\bar{a},\bar{c}\right)$ when $\bar{a} \subseteq \bar{b} \subseteq \bar{c}$.*

Proof: When $\bar{a} \subseteq \bar{b} \subseteq \bar{c}$, then

$$S_{S_h}\left(\bar{a},\bar{b}\right) = \bigstar \left(\frac{|\bar{a} \cap \bar{b}|}{|\bar{a}|}, \frac{|\bar{a} \cap \bar{b}|}{|\bar{b}|} \right) = \bigstar \left(\frac{|\bar{a}|}{|\bar{a}|}, \frac{|\bar{a}|}{|\bar{b}|} \right) = \frac{|\bar{a}|}{|\bar{b}|},$$

$$S_{S_h}\left(\bar{a},\bar{c}\right) = \bigstar \left(\frac{|\bar{a} \cap \bar{c}|}{|\bar{a}|}, \frac{|\bar{a} \cap \bar{c}|}{|\bar{c}|} \right) = \bigstar \left(\frac{|\bar{a}|}{|\bar{a}|}, \frac{|\bar{a}|}{|\bar{c}|} \right) = \frac{|\bar{a}|}{|\bar{c}|}.$$

Thus,

$$D_{S_h}\left(\bar{a},\bar{b}\right) = 1 - S_{S_h}\left(\bar{a},\bar{b}\right) = 1 - \frac{|\bar{a}|}{|\bar{b}|},$$

$$D_{S_h}\left(\bar{a},\bar{c}\right) = 1 - S_{S_h}\left(\bar{a},\bar{c}\right) = 1 - \frac{|\bar{a}|}{|\bar{c}|}.$$

As $\bar{b} \subseteq \bar{c}$, it follows that $|\bar{b}| \le |\bar{c}|$. Therefore, $\frac{|\bar{a}|}{|\bar{b}|} \ge \frac{|\bar{a}|}{|\bar{c}|}$ which implies that

$$1 - \frac{|\bar{a}|}{|\bar{b}|} \le 1 - \frac{|\bar{a}|}{|\bar{c}|}.$$

Hence $D_{S_h}\left(\bar{a},\bar{b}\right) \le D_{S_h}\left(\bar{a},\bar{c}\right)$. ∎
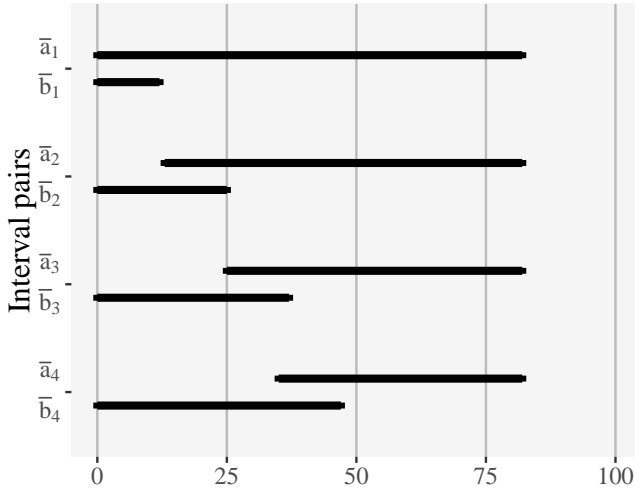
Fig. 3. Interval pairs used to demonstrate the *aliasing* of SMs

TABLE III
THE S_J AND S_D MEASURES EXHIBIT ALIASING FOR THE INTERVAL
PAIRS IN FIG. 3

Interval Pair	S_J	S_D	$S_{S_h \wedge}$	$S_{S_h \Pi}$
$\{\bar{a}_1, \bar{b}_1\}$	0.15	0.26	0.15	0.15
$\{\bar{a}_2, \bar{b}_2\}$	0.15	0.26	0.17	0.08
$\{\bar{a}_3, \bar{b}_3\}$	0.15	0.26	0.21	0.07
$\{\bar{a}_4, \bar{b}_4\}$	0.15	0.26	0.26	0.06

IV. DEMONSTRATION AND ANALYSIS

We now demonstrate and analyse the behaviour of the proposed SM S_{S_h} with the minimum ($\wedge$) and product (Π) *t-norm*s in the context of the S_J and S_D SMs for a set of synthetic examples in the first part and with a real dataset in the last part. Here, we alter different features of interval-pairs to explore how well these three measures perform or follow intuitive results. In particular, we focus on the following aspects:

1) Propensity to exhibiting *aliasing* in response to variations in interval sizes.
2) Behaviour in respect to intervals where one is a complete subset of the other.
3) Behaviour in respect to intervals of equal sizes and overlapping ratio.
4) Response to variations in interval size, while maintaining the same level of subsethood.
5) Response to linear increase in the overlap of intervals.

### A. Synthetic Dataset Based Demonstration

For each of the above cases, a series of synthetic intervals is proposed and visualised.

1) Experiment on aliasing propensity: In Fig. 3, four different pairs of intervals $\{\bar{a}, \bar{b}\}$ are considered where all pairs have an intersection of equal size. The similarity results for the pairs using the three SMs are shown in Table III. The S_J and S_D measures are subject to *aliasing*, providing the same similarity of 0.15 and 0.26 respectively for all pairs.

Indeed, both measures provide –unexpectedly– identical similarities for pairs of intervals when the size of the union
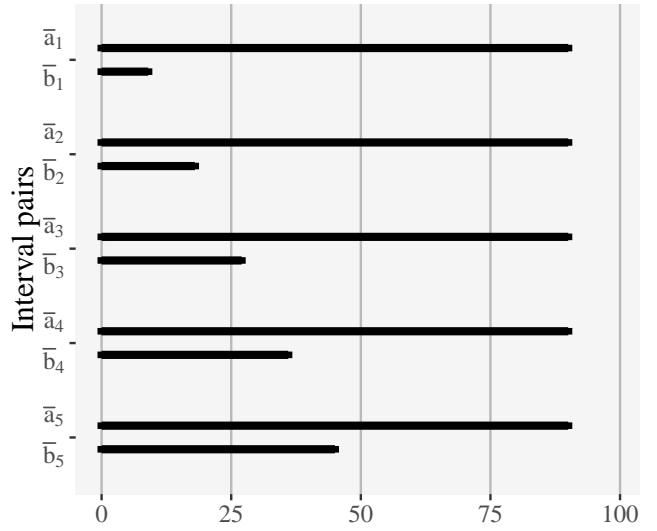
(a) Increasing interval overlapping
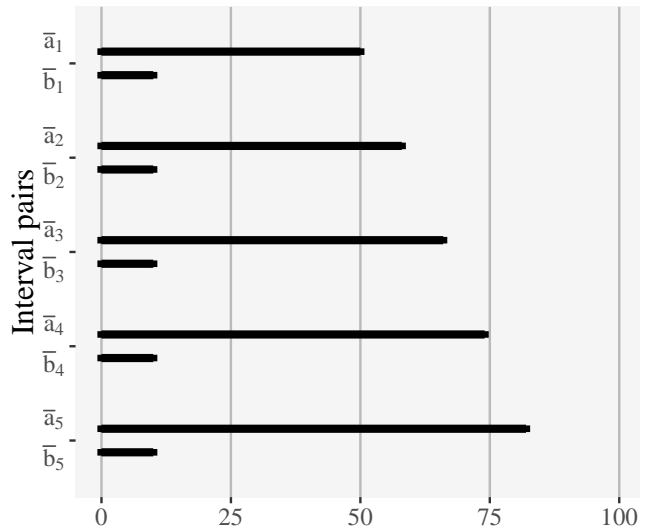
(b) Increasing interval non-overlapping

Fig. 4. Interval pairs where one interval as a complete subset of the other

TABLE IV
SIMILARITY RESULTS FOR THE INTERVAL PAIRS AS SHOWN IN FIG. 4

Interval Pair	S_J	S_D	$S_{S_h \wedge}$	$S_{S_h \Pi}$	subset by
$\{\bar{a}_1, \bar{b}_1\}$	0.10	0.18	0.10	0.10	10%
$\{\bar{a}_2, \bar{b}_2\}$	0.20	0.33	0.20	0.20	20%
$\{\bar{a}_3, \bar{b}_3\}$	0.30	0.46	0.30	0.30	30%
$\{\bar{a}_4, \bar{b}_4\}$	0.40	0.57	0.40	0.40	40%
$\{\bar{a}_5, \bar{b}_5\}$	0.50	0.67	0.50	0.50	50%

(a) Increasing interval overlapping

Interval Pair	S_J	S_D	$S_{S_h \wedge}$	$S_{S_h \Pi}$	subset by
$\{\bar{a}_1, \bar{b}_1\}$	0.20	0.33	0.20	0.20	20%
$\{\bar{a}_2, \bar{b}_2\}$	0.172	0.294	0.172	0.172	17.2%
$\{\bar{a}_3, \bar{b}_3\}$	0.152	0.263	0.152	0.152	15.2%
$\{\bar{a}_4, \bar{b}_4\}$	0.135	0.238	0.135	0.135	13.5%
$\{\bar{a}_5, \bar{b}_5\}$	0.122	0.217	0.122	0.122	12.2%

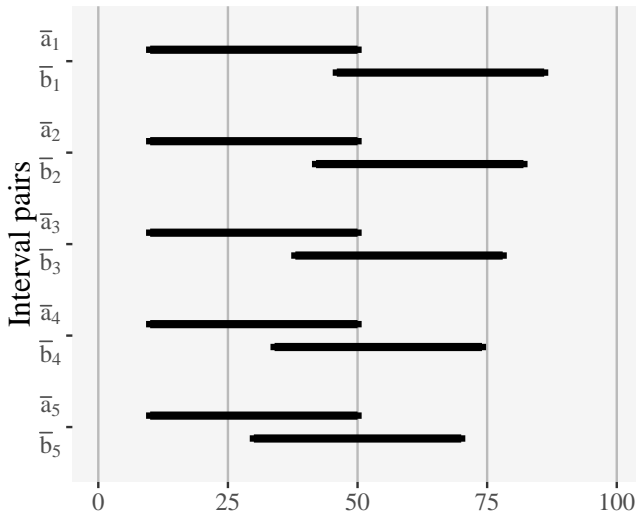(b) Increasing interval non-overlapping

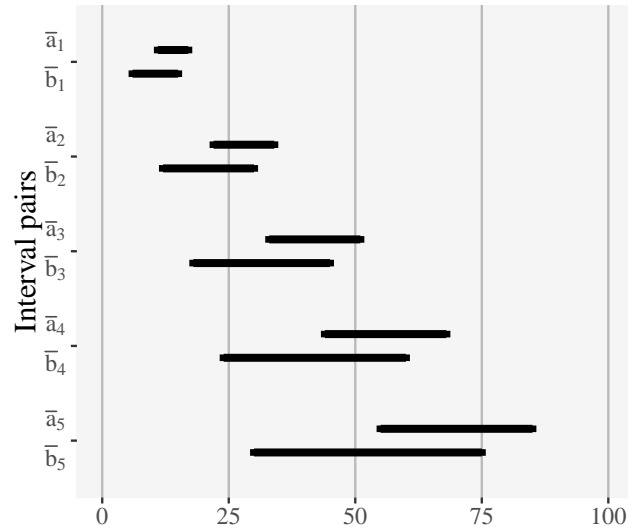Fig. 5. Interval pairs with equal size and equal ratio of overlap

Fig. 6. Interval pairs used to demonstrate the invariance of SMs

TABLE V
SIMILARITY RESULTS FOR THE INTERVAL PAIRS AS SHOWN IN FIG. 5

Interval Pair	S_J	S_D	$S_{S_h \wedge}$	$S_{S_h \Pi}$	intersecting by
$\{\overline{a}_1, \overline{b}_1\}$	0.05	0.10	0.10	0.01	10%
$\{\overline{a}_2, \overline{b}_2\}$	0.11	0.20	0.20	0.04	20%
$\{\overline{a}_3, \overline{b}_3\}$	0.18	0.30	0.30	0.09	30%
$\{\overline{a}_4, \overline{b}_4\}$	0.25	0.40	0.40	0.16	40%
$\{\overline{a}_5, \overline{b}_5\}$	0.33	0.50	0.50	0.25	50%

TABLE VI
ALL THREE SMS DISPLAY INVARIANT SIMILARITY FOR THE INTERVAL
PAIRS IN FIG. 6

Interval Pair	S_J	S_D	$S_{S_h \wedge}$	$S_{S_h \Pi}$	multiplied by n
$\{\overline{a}_1, \overline{b}_1\}$	0.36	0.53	0.44	0.29	1
$\{\overline{a}_2, \overline{b}_2\}$	0.36	0.53	0.44	0.29	2
$\{\overline{a}_3, \overline{b}_3\}$	0.36	0.53	0.44	0.29	3
$\{\overline{a}_4, \overline{b}_4\}$	0.36	0.53	0.44	0.29	4
$\{\overline{a}_5, \overline{b}_5\}$	0.36	0.53	0.44	0.29	5

of their non-overlapping segments remains constant. On the contrary, the S_{S_h} measure with both minimum and product *t-norms* ($S_{S_h \wedge}$ and $S_{S_h \Pi}$) yields a different degree of similarity for all cases, thereby exhibiting its *aliasing*-free and robust behaviour. The reason is that the S_{S_h} measure captures the changes in the size of both input intervals as compared to S_J and S_D measures, which eventually affects their reciprocal subsethood and the overall similarity. Note that as shown in Theorem 11, the results of the $S_{S_h \wedge}$ measure are bounded by the S_J and S_D measures.

2) Experiment with interval pairs when one interval is a complete subset of the other: Two separate cases are considered with different sets of interval pairs, where in both cases one interval is a complete subset of the other: (a) an increasing degree of overlap between both intervals by increasing the size of the smaller interval, and (b) decreasing the degree of overlap by increasing the size of the larger interval.

Figure 4(a) shows five interval pairs with $\overline{b} \subset \overline{a}$, where $\overline{b}$ covers 10%, 20%, 30%, 40%, and 50% of $\overline{a}$. On the contrary, in Fig. 4(b) another five interval pairs are presented where the degree of overlap (in %) is gradually decreased by increasing the size of $\overline{a}$. Tables IV(a) and IV(b) show the similarity for all pairs with all three SMs. Note that for all pairs, the subsethood of $\overline{b}$ is 1 while the subsethood of $\overline{a}$ depends on the size of $\overline{a}$ and $\overline{b}$, i.e., $\frac{|\overline{a} \cap \overline{b}|}{|\overline{a}|}$. Intuitively, their mutual similarity should be at most $\frac{|\overline{a} \cap \overline{b}|}{|\overline{a}|}$ for each pair. From the results, we see that both S_{S_h} (with minimum and product *t-norms*) and S_J measures perform according to the intuition while the S_D measure exceeds this expected limit.

3) Experiment with interval pairs of equal size and equal overlapping ratio: In Fig. 5, five interval pairs are shown where the intervals are of equal size and the size of their intersection is varied to 10%, 20%, 30%, 40%, and 50% of their size. Table V provides the results for all pairs using the three SMs. In all pairs, the subsethood is equal, and it is intuitive to expect the similarity to be the same as this subsethood (as the intervals are both of equal size). In this case, the S_{S_h} with minimum *t-norm* ($S_{S_h \wedge}$) and S_D measures follow the intuition whereas the S_{S_h} with product *t-norm* ($S_{S_h \Pi}$) and S_J measures yield a lower degree of similarity.

4) Experiment on scaling invariance: Five pairs of intervals are shown in Fig. 6, where both endpoints of $\overline{a}$ and $\overline{b}$ are gradually multiplied by a factor $n \in \{2, 3, 4, 5\}$ to generate new interval pairs. The degree of subsethood stays the same across the pairs. Adapting the definition from [54], a similarity measure is *invariant* if its similarity output remains constant regardless of scaling the interval endpoints by a factor. Table VI shows the similarity for all pairs using the three SMs where n is the factor applied to the interval endpoints. The results demonstrate the *scaling invariance* property for the given pairs of intervals for all SMs.

5) Experiment on increased overlap linearly: Adapting the definition from [54], a SM on intervals is *linear* if its similarity output varies linearly as to a linear change in the size of the overlap/intersection of the intervals. In Fig. 7(a), the overlap between two intervals of equal size is gradually increased in 10% steps. The corresponding similarity outputs for the pairs
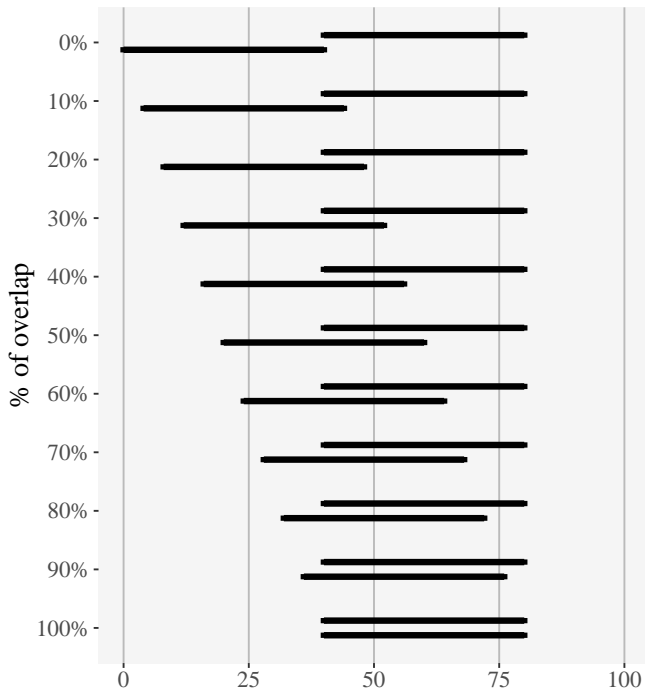
(a) Rising overlap while intervals are of equal size
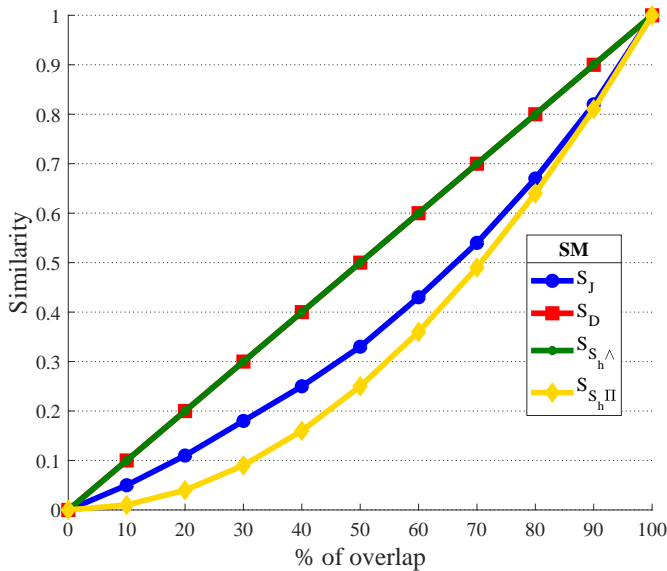
(b) Similarity versus overlap

Fig. 7. Displaying linearity by the S_D and $S_{S_h \wedge}$ measures, and convexity by the S_J and $S_{S_h \Pi}$ measures

and all three SMs are shown graphically in Fig. 7(b). Results show that all three SMs provide higher similarity for increased overlap. Particularly, the S_{S_h} with minimum *t-norm* ($S_{S_h \wedge}$) and S_D measures exhibit *linearity* in the similarity results while the S_{S_h} with product *t-norm* ($S_{S_h \Pi}$) and S_J measure display *convexity* (similarity increases rapidly with an increase in the size of the intersection).

In summary, case 1) shows that S_J and S_D measures are subject to *aliasing* and return the same similarity for very different interval-pairs whenever the union of their non-overlapping remains constant (regardless of any changes in

their size). In case 2), when one interval is a complete subset of the other, proportionately increasing or decreasing their degree of overlap results in the S_{S_h} and S_J measures returning sensible results whereas the S_D measure overestimates the similarity. In case 3), the S_{S_h} with minimum *t-norm* ($S_{S_h \wedge}$) and S_D measures meet the expectation when intervals are of equal size and their intersection is changed proportionately, whereas the S_J measure underestimates the similarity. In case 4), we multiply interval end-points by a factor, maintaining the same degree of subsethood. Here, all three measures show *scaling invariance* in the results as expected. Lastly, in case 5), for linear increases in the size of overlap between the intervals of equal size, the S_{S_h} with minimum *t-norm* ($S_{S_h \wedge}$) and S_D measures exhibit *linearity* in the similarity results while the S_{S_h} with product *t-norm* ($S_{S_h \Pi}$) and S_J measure display *convexity*. These experiments demonstrate that the proposed SM shows behaviour in line with expectation, which is not the case for the other, commonly used measures.

B. A Real World Example

In this part, we have used a real world dataset to review the behaviour of the three SMs (S_{S_h} with minimum ($\wedge$) and product (Π) *t-norm*s, S_J, and S_D), particularly in respect to the *aliasing* issue. The dataset used for this demonstration is the temperature data of different standard areas (districts) of the UK for the year 2016 from the UK Met Office. The entire dataset is available at [55]. The map of the standard areas used by the UK Met office is shown in Fig. 8. We have used temperature data of nine standard areas: Scotland North (SN), Scotland East (SE), Scotland West (SW), England East & North East (EEN), England Northwest & Wales North (ENW), Midlands (ML), East Anglia (EA), England Southwest & Wales South (ESW), and England South-east & Central South (ESC) for four different seasons in 2016.

The intervals are constructed by taking the minimum and the maximum temperature of each area for every season, as shown in Table VII. We have applied the S_{S_h} (with minimum ($\wedge$) and product (Π) *t-norm*s), S_J and S_D measures to estimate similarities between all possible pairs of the nine areas for all seasons; here the results for Autumn and Spring seasons are reported which are shown in Tables VIII, IX, X and XI. We do not include results for other two seasons—Winter and Summer—as they follow the same pattern.

Regarding similarity based on seasonal temperature, one can expect higher similarity for pairs of areas geographically close to each other (i.e., located in the same latitude), while expecting lower similarity for remote areas as it is generally recognized that latitude (i.e., distance from the equator) is an important factor affecting variations in temperature. Hence, in this real world example, the districts in the Northern UK (e.g., SE, SW) are likely to be more similar as they are in the South (e.g., ML, EA). By the same token, a district in the North will be highly dissimilar to a district in the South. Generally, all three SMs appear to meet up with such expectations. For instance, considering Autumn-temperatures between SN and the other eight areas from geographically close to far off, three SMs appear to generate higher to lower similarity as temperature-range differs more.
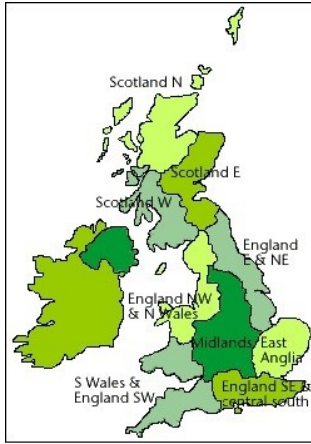
Fig. 8. Standard areas (districts) of the UK [55]

TABLE VII
SEASONAL TEMPERATURE IN STANDARD AREAS OF THE UK IN 2016

Area	Acronym	Seasons							
		Winter		Spring		Summer		Autumn	
		Min	Max	Min	Max	Min	Max	Min	Max
Scotland N	SN	0.5	6.0	2.7	9.8	9.1	15.8	4.8	11.2
Scotland E	SE	0.4	6.2	2.4	10.3	9.1	16.9	4.4	11.5
Scotland W	SW	1.9	7.2	3.2	11.1	10.0	17.3	5.6	12.0
England E & NE	EEN	2.4	8.5	3.6	11.6	10.9	19.5	6.6	13.5
England NW & Wales N	ENW	3.1	8.6	3.9	11.8	11.4	18.7	6.5	13.1
Midlands	ML	3.3	9.4	3.9	12.6	11.7	20.6	6.9	14.1
East Anglia	EA	3.8	10.0	4.5	13.2	12.3	22.1	7.7	15.5
England SW & Wales S	ESW	4.5	9.8	4.4	12.7	11.9	19.5	7.3	14.2
England SE & Central S	ESC	4.3	10.2	4.7	13.5	12.4	21.6	7.5	15.4

TABLE VIII
SIMILARITY RESULTS FOR TEMPERATURE DATA USING THE S_J MEASURE

Area	Autumn									Spring								
	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC
SN	1	0.9014	0.7778	0.5287	0.5663	0.4624	0.3271	0.4149	0.3491	1	0.8987	0.7857	0.6966	0.6484	0.596	0.5048	0.54	0.4722
SE	0.9014	1	0.7763	0.5385	0.5747	0.4742	0.3423	0.4286	0.3636	0.8987	1	0.8161	0.7283	0.6809	0.6275	0.537	0.5728	0.5045
SW	0.7778	0.7763	1	0.6835	0.7333	0.6	0.4343	0.5465	0.4592	0.7857	0.8161	1	0.8929	0.8372	0.766	0.66	0.7053	0.6214
EEN	0.5287	0.5385	0.6835	1	0.9286	0.88	0.6517	0.8158	0.6818	0.6966	0.7283	0.8929	1	0.939	0.8556	0.7396	0.7912	0.697
ENW	0.5663	0.5747	0.7333	0.9286	1	0.8158	0.6	0.7532	0.6292	0.6484	0.6809	0.8372	0.939	1	0.908	0.7849	0.8409	0.7396
ML	0.4624	0.4742	0.6	0.88	0.8158	1	0.7442	0.9315	0.7765	0.596	0.6275	0.766	0.8556	0.908	1	0.871	0.9318	0.8229
EA	0.3271	0.3423	0.4343	0.6517	0.6	0.7442	1	0.7927	0.9625	0.5048	0.537	0.66	0.7396	0.7849	0.871	1	0.9318	0.9444
ESW	0.4149	0.4286	0.5465	0.8158	0.7532	0.9315	0.7927	1	0.8272	0.54	0.5728	0.7053	0.7912	0.8409	0.9318	0.9318	1	0.8791
ESC	0.3491	0.3636	0.4592	0.6818	0.6292	0.7765	0.9625	0.8272	1	0.4722	0.5045	0.6214	0.697	0.7396	0.8229	0.9444	0.8791	1

TABLE IX
SIMILARITY RESULTS FOR TEMPERATURE DATA USING THE S_D MEASURE

Area	Autumn									Spring								
	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC
SN	1	0.9481	0.875	0.6917	0.7231	0.6324	0.493	0.5865	0.5175	1	0.9467	0.88	0.8212	0.7867	0.7468	0.6709	0.7013	0.6415
SE	0.9481	1	0.8741	0.7	0.7299	0.6434	0.5101	0.6	0.5333	0.9467	1	0.8987	0.8428	0.8101	0.7711	0.6988	0.7284	6707
SW	0.875	0.8741	1	0.812	0.8462	0.75	0.6056	0.7068	0.6294	0.88	0.8987	1	0.9434	0.9114	0.8675	0.7952	0.8272	0.7665
EEN	0.6917	0.7	0.812	1	0.963	0.9362	0.7891	0.8986	0.8108	0.8212	0.8428	0.9434	1	0.9686	0.9222	0.8503	0.8834	0.8214
ENW	0.7231	0.7299	0.8462	0.963	1	0.8986	0.75	0.8593	0.7724	0.7867	0.8101	0.9114	0.9686	1	0.9518	0.8795	0.9136	0.8503
ML	0.6324	0.6434	0.75	0.9362	0.8986	1	0.8533	0.9645	0.8742	0.7468	0.7711	0.8675	0.9222	0.9518	1	0.931	0.9647	0.9029
EA	0.493	0.5101	0.6056	0.7891	0.75	0.8533	1	0.8844	0.9809	0.6709	0.6988	0.7952	0.8503	0.8795	0.931	1	0.9647	0.9714
ESW	0.5865	0.6	0.7068	0.8986	0.8593	0.9645	0.8844	1	0.9054	0.7013	0.7284	0.8272	0.8834	0.9136	0.9647	0.9647	1	0.9357
ESC	0.5175	0.5333	0.6294	0.8108	0.7724	0.8742	0.9809	0.9054	1	0.6415	0.6707	0.7665	0.8214	0.8503	0.9029	0.9714	0.9357	1

TABLE X
SIMILARITY RESULTS FOR TEMPERATURE DATA USING THE $S_{S_h \wedge}$ MEASURE

Area	Autumn									Spring								
	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC
SN	1	0.9014	0.875	0.6667	0.7121	0.5972	0.4487	0.5652	0.4684	1	0.8987	0.8354	0.775	0.7468	0.6782	0.6092	0.6506	0.5795
SE	0.9014	1	0.831	0.6901	0.7042	0.6389	0.4872	0.5915	0.5063	0.8987	1	0.8987	0.8375	0.8101	0.7356	0.6667	0.7108	0.6364
SW	0.875	0.831	1	0.7826	0.8333	0.7083	0.5513	0.6812	0.5696	0.8354	0.8987	1	0.9375	0.9114	0.8276	0.7586	0.8072	0.7273
EEN	0.6667	0.6901	07826	1	0.942	0.9167	0.7436	0.8986	0.7595	0.775	0.8375	0.9375	1	0.9625	0.8851	0.8161	0.8675	0.7841
ENW	0.7121	0.7042	0.8333	0.942	1	0.8611	0.6923	0.8406	0.7089	0.7468	0.8101	0.9114	0.9625	1	0.908	0.8391	0.8916	0.8068
ML	0.5972	0.6389	0.7083	0.9167	0.8611	1	0.8205	0.9444	0.8354	0.6782	0.7356	0.8276	0.8851	0.908	1	0.931	0.9425	0.8977
EA	0.4487	0.4872	0.5513	0.7436	0.6923	0.8205	1	0.8333	0.9747	0.6092	0.6667	0.7586	0.8161	0.8391	0.931	1	0.9425	0.9659
ESW	0.5652	0.5915	0.6812	0.8986	0.8406	0.9444	0.8333	1	0.8481	0.6506	0.7108	0.8072	0.8675	0.8916	0.9425	0.9425	1	0.9091
ESC	0.4684	0.5063	0.5696	0.7595	0.7089	0.8354	0.9747	0.8481	1	0.5795	0.6364	0.7273	0.7841	0.8086	0.8977	0.9659	0.9091	1

TABLE XI
SIMILARITY RESULTS FOR TEMPERATURE DATA USING THE $S_{S_h\Pi}$ MEASURE

Area	Autumn									Spring								
	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC
SN	1	0.9014	0.7656	0.4792	0.523	0.4013	0.2454	0.3444	0.2708	1	0.8987	0.7766	0.6768	0.6206	0.5635	0.4548	0.4948	0.4163
SE	0.9014	1	0.7661	0.4901	0.5335	0.4139	0.2607	0.3601	0.2853	0.8987	1	0.8077	0.7103	0.6563	0.596	0.4895	0.5309	0.4511
SW	0.7656	0.7661	1	0.6603	0.7161	0.5645	0.3704	0.5002	0.4005	0.7766	0.8077	1	0.89	0.8306	0.7543	0.6338	0.6846	0.5892
EEN	0.4792	0.4901	0.6603	1	0.9278	0.8768	0.625	0.8074	0.6604	0.6768	0.7103	0.89	1	0.9381	0.8519	0.7243	0.7807	0.6763
ENW	0.523	0.5335	0.7161	0.9278	1	0.8089	0.5664	0.7387	0.6015	0.6206	0.6563	0.8306	0.9381	1	0.908	0.7754	0.8351	0.7251
ML	0.4013	0.4139	0.5645	0.8768	0.8089	1	0.7293	0.9308	0.7658	0.5635	0.596	0.7543	0.8519	0.908	1	0.8668	0.9312	0.8152
EA	0.2454	0.2607	0.3704	0.625	0.5664	0.7293	1	0.785	0.9622	0.4548	0.4895	0.6338	0.7243	0.7754	0.8668	1	0.9312	0.9437
ESW	0.3444	0.3601	0.5002	0.8074	0.7387	0.9308	0.785	1	0.8235	0.4948	0.5309	0.6846	0.7807	0.8351	0.9312	0.9312	1	0.8762
ESC	0.2708	0.2853	0.4005	0.6604	0.6015	0.7658	0.9622	0.8235	1	0.4163	0.4511	0.5892	0.6763	0.7251	0.8152	0.9437	0.8762	1

However, in several cases, the S_J and S_D measures suffer from the *aliasing*—returning the same similarity for different ranges of temperature. As mentioned earlier, both measures provide identical similarity for pairs of areas as long as the size of the intersection and union of their temperature-range remain constant. Table VIII and IX show that they produce the same similarity (i.e., 0.7396 and 0.8503 respectively) for pairs of areas, (EEN versus EA) and (ENW versus ESC) in Spring, even though they have separate temperature-ranges. Analogously, we also notice identical similarities for the pairs, (SW vs ML) and (EA vs ENW), and the pairs (ML vs ENW) and (ESW vs EEN) in Autumn. In contrast, the S_{S_h} measure with minimum and product *t-norm*s (i.e., $S_{S_h\wedge}$ and $S_{S_h\Pi}$), responding to variation in temperature-range, yields a distinct outcome for all of these pairs as shown in Table X and XI.

This real world example showing the *aliasing* inherent to the S_J and S_D measures highlights the potential for misleading inference in practice. In particular, while clustering the areas with respect to temperature, some areas may be placed in the same group albeit having different ranges of temperature. This can be an issue when we compare real-world interval data for the purpose of grouping, ranking, and decision making. Therefore, a similarity measure which avoids this *aliasing*—such as the proposed measure, is desirable.

V. CONCLUSIONS AND FUTURE WORK

The contributions of this paper centre on to identification and articulation of potential shortcomings – specifically *aliasing* – affecting popular set-theoretic similarity measures. The latter have the potential to give misleading results, in turn affecting real-world applications dependent on the robust comparison of intervals. In order to address this limitation, this paper develops the underpinnings of robust SMs by putting forward five new axioms which complement the traditional axioms associated with the most common SMs. Building on this set of nine axioms, the paper establishes a new set-theoretic similarity measure for intervals based on their bidirectional subsethood. The new SM is shown to avoid shortcomings such as *aliasing*, while delivering intuitive results (e.g., being bounded above and below by Jaccard and Dice SMs), facilitating its potential use in real world applications. As part of the development of the new SM, the paper also puts forward the definition of subsethood for intervals and proofs of its mathematical properties (see Appendix). Finally,

a corresponding dissimilarity or distance measure is derived from the new SM, which is also proven to be a metric.

At an experimental level, the paper provides a detailed investigation contrasting the behaviour of the proposed SM vis-a-vis the Jaccard and Dice SMs – using both synthetic and real-world interval-valued data. The exhaustive analyses confirm that the new measure exhibits desirable behaviour while maintaining all essential features of a similarity measure. In particular, the new SM is resilient to *aliasing* and provides desirable results in respect to all key features (e.g., *linearity*, *scaling invariance*), whereas popular SMs are shown to produce counter-intuitive results in some cases.

In the future, we plan to use this new measure for assessing similarity between discontinuous intervals and to develop corresponding extensions of the SM for type-1, interval-valued, and type-2 fuzzy sets using the α-cut decomposition representation. We also aim to explore it for capturing the mutual agreement of interval-valued evidence for aggregation as well as clustering and classification of real-world interval-valued data sets.

APPENDIX

DEFINITION OF SUBSETHOOD FOR CLOSED INTERVALS

Building on the axiomatic definitions of subsethood in the literature (see Section II.C), the key properties of subsethood for intervals are captured in the following theorem.

Theorem 14. *A real-valued function $S_h : \overline{a} \times \overline{b} \to [0,1]$ captures subsethood for intervals, if it follows the properties,*
(1) $S_h(\overline{a},\overline{b}) = 1$ if and only if $\overline{a} \subseteq \overline{b}$,
(2) $S_h(\overline{a},\overline{b}) = 0$ if and only if $\overline{a}$ and $\overline{b}$ are disjoint,
(3) if $\overline{a} \subseteq \overline{b} \subseteq \overline{c}$, then $S_h(\overline{c},\overline{a}) \leq S_h(\overline{b},\overline{a})$ and $S_h(\overline{c},\overline{a}) \leq S_h(\overline{c},\overline{b})$.

Proof: Below we give proofs for all the properties (1–3).
(1) If $\overline{a} \subseteq \overline{b}$, then $|\overline{a} \cap \overline{b}| = |\overline{a}|$. Thus, $S_h(\overline{a},\overline{b}) = \frac{|\overline{a}\cap\overline{b}|}{|\overline{a}|} = \frac{|\overline{a}|}{|\overline{a}|} = 1$. Again, $S_h(\overline{a},\overline{b}) = 1$ only when $\overline{a}$ is included in $\overline{b}$.
(2) If $\overline{a}$ and $\overline{b}$ are disjoint, then $|\overline{a}\cap\overline{b}| = 0$. Thus, $S_h(\overline{a},\overline{b}) = 0$. Contrarily, $S_h(\overline{a},\overline{b}) = 0$ when $\overline{a}$ and $\overline{b}$ are not overlapping.
(3) If $\overline{a} \subseteq \overline{b} \subseteq \overline{c}$, then $|\overline{a}| \leq |\overline{b}| \leq |\overline{c}|$. Thus, $S_h(\overline{c},\overline{a}) = \frac{|\overline{c}\cap\overline{a}|}{|\overline{c}|} = \frac{|\overline{a}|}{|\overline{c}|}$ and $S_h(\overline{b},\overline{a}) = \frac{|\overline{b}\cap\overline{a}|}{|\overline{b}|} = \frac{|\overline{a}|}{|\overline{b}|}$. Since $|\overline{b}| \leq |\overline{c}|$, it follows that $\frac{1}{|\overline{c}|} \leq \frac{1}{|\overline{b}|}$. Hence, $\frac{|\overline{a}|}{|\overline{c}|} \leq \frac{|\overline{a}|}{|\overline{b}|}$, implying $S_h(\overline{c},\overline{a}) \leq S_h(\overline{b},\overline{a})$. Moreover, $S_h(\overline{c},\overline{b}) = \frac{|\overline{c}\cap\overline{b}|}{|\overline{c}|} = \frac{|\overline{b}|}{|\overline{c}|}$. Since $|\overline{a}| \leq |\overline{b}|$, it follows that $\frac{|\overline{a}|}{|\overline{c}|} \leq \frac{|\overline{b}|}{|\overline{c}|}$, proving that $S_h(\overline{c},\overline{a}) \leq S_h(\overline{c},\overline{b})$. ∎

References

[1] A. Tversky, "Features of similarity," *Psychological Review*, vol. 84, no. 4, pp. 327–352, 1977.

[2] I. Beg and S. Ashraf, "Similarity measures for fuzzy sets," *Applied and Computational Mathematics*, vol. 8, no. 2, pp. 192–202, 2009.

[3] M. Galar, J. Fernández, G. Beliakov, and H. Bustince, "Interval-valued fuzzy sets applied to stereo matching of color images," *IEEE Trans. Image Processing*, vol. 20, no. 7, pp. 1949–1961, 2011.

[4] C. Wagner, S. Miller, J. M. Garibaldi, D. T. Anderson, and T. C. Havens, "From interval-valued data to general type-2 fuzzy sets," *IEEE Trans. Fuzzy Systems*, vol. 23, no. 2, pp. 248–269, 2015.

[5] D. S. Guru, B. B. Kiranagi, and P. Nagabhushan, "Multivalued type proximity measure and concept of mutual similarity value useful for clustering symbolic patterns," *Pattern Recognition Letters*, vol. 25, no. 10, pp. 1203–1213, 2004.

[6] Y. Ren, Y.-H. Liu, J. Rong, and R. Dew, "Clustering interval-valued data using an overlapped interval divergence," in *Proc. 8th Australasian Data Mining Conf.*, 2009, pp. 35–42.

[7] P. Jaccard, "Nouvelles recherches sur la distribution florale," *Bulletin de la Socit vaudoise des Sciences Naturelles*, vol. 44, pp. 223–270, 1908.

[8] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[9] S. Kabir, C. Wagner, T. C. Havens, D. T. Anderson, and U. Aickelin, "Novel similarity measure for interval-valued data based on overlapping ratio," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2017, pp. 1–6.

[10] D. Szymkiewicz, "Une conlribution statistique à la géographie floristique," *Acta Societatis Botanicorum Poloniae*, vol. 11, no. 3, pp. 249–265, 1934.

[11] Y. Otsuka, "The faunal character of the japanese pleistocene marine mollusca, as evidence of climate having become colder during the pleistocene in japan," *Bulletin of the Biogeographical Society of Japan*, vol. 6, no. 16, pp. 165–170, 1936.

[12] R. R. Sokal and P. H. A. Sneath, "Principles of numerical taxonomy." *Principles of numerical taxonomy*, 1963.

[13] R. R. Sokal and C. D. Michener, "A statistical method for evaluating systematic relationship," *University of Kansas science bulletin*, vol. 28, pp. 1409–1438, 1958.

[14] J. McCulloch, C. Wagner, and U. Aickelin, "Analysing fuzzy sets through combining measures of similarity and distance," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2014, pp. 155–162.

[15] K. H. Rosen, *Discrete Mathematics and Its Applications*, 7th ed. McGraw-Hill, 2012.

[16] T. C. Havens, D. T. Anderson, C. Wagner, H. Deilamsalehy, and D. Wonnacott, "Fuzzy integrals of crowd-sourced intervals using a measure of generalized accord," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2013, pp. 1–8.

[17] T. C. Havens, D. T. Anderson, and C. Wagner, "Data-informed fuzzy measures for fuzzy integration of intervals and fuzzy numbers," *IEEE Trans. Fuzzy Systems*, vol. 23, no. 5, pp. 1861–1875, 2015.

[18] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. Academic press, 1980.

[19] C. Wagner, D. T. Anderson, and T. C. Havens, "Generalization of the fuzzy integral for discontinuous interval-and non-convex interval fuzzy set-valued inputs," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2013, pp. 1–8.

[20] G. Beliakov and S. James, "A penalty-based aggregation operator for non-convex intervals," *Knowledge-Based Systems*, vol. 70, pp. 335–344, 2014.

[21] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[22] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Upper Saddle River, NJ: Prentic Hall PTR, 2001.

[23] C. Wagner, S. Miller, and J. M. Garibaldi, "Similarity based applications for data-driven concept and word models based on type-1 and type-2 fuzzy sets," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2013, pp. 1–9.

[24] H. T. Nguyen and K. Vladik, "Computing degrees of subsethood and similarity for interval-valued fuzzy sets: fast algorithms," in *Proc. 9th Int. Conf. Intelligent Technologies*, 2008, pp. 47–55.

[25] A. Stachowiak, P. Żywica, K. Dyczkowski, and A. Wójtowicz, "An interval-valued fuzzy classifier based on an uncertainty-aware similarity measure," in *Intelligent Systems' 2014*. Springer, pp. 741–751.

[26] D. Wu and J. M. Mendel, "A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets," *Information Sciences*, vol. 179, no. 8, pp. 1169–1192, 2009.

[27] J. McCulloch, C. Wagner, and U. Aickelin, "Extending similarity measures of interval type-2 fuzzy sets to general type-2 fuzzy sets," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2013, pp. 1–8.

[28] J. Ye, "Multicriteria decision-making method using the dice similarity measure between expected intervals of trapezoidal fuzzy numbers," *J. Decision Systems*, vol. 21, no. 4, pp. 307–317, 2012.

[29] ——, "The dice similarity measure between generalized trapezoidal fuzzy numbers based on the expected interval and its multicriteria group decision-making method," *J. Chinese Institute of Industrial Engineers*, vol. 29, no. 6, pp. 375–382, 2012.

[30] L. Baccour and R. I. John, "Experimental analysis of crisp similarity and distance measures," in *Proc. 6th Int. Con. Soft Computing and Pattern Recognition*, 2014, pp. 96–100.

[31] A. Gardner, J. Kanno, C. A. Duncan, and R. Selmic, "Measuring distance between unordered sets of different sizes," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2014, pp. 137–143.

[32] Y. Li, K. Qin, and X. He, "Relations among similarity measure, subsethood measure and fuzzy entropy," *Int. J. Computational Intelligence Systems*, vol. 6, no. 3, pp. 411–422, 2013.

[33] B. Kosko, "Fuzzy entropy and conditioning," *Information Sciences*, vol. 40, no. 2, pp. 165–174, 1986.

[34] W. Bandler and L. Kohout, "Fuzzy power sets and fuzzy implication operators," *Fuzzy Sets and Systems*, vol. 4, no. 1, pp. 13–30, 1980.

[35] L. Kitainik, "Fuzzy inclusions and fuzzy dichotomous decision procedures," in *Optimization Models Using Fuzzy Sets and Possibility Theory*. Springer, 1987, pp. 154–170.

[36] D. Sinha and E. R. Dougherty, "Fuzzification of set inclusion: theory and applications," *Fuzzy Sets and Systems*, vol. 55, no. 1, pp. 15–42, 1993.

[37] V. R. Young, "Fuzzy subsethood," *Fuzzy Sets and Systems*, vol. 77, no. 3, pp. 371–384, 1996.

[38] J. Fan, W. Xie, and J. Pei, "Subsethood measure: new definitions," *Fuzzy Sets and Systems*, vol. 106, no. 2, pp. 201–209, 1999.

[39] H. Bustince, V. Mohedano, E. Barrenechea, and M. Pagola, "Definition and construction of fuzzy di-subsethood measures," *Information Sciences*, vol. 176, no. 21, pp. 3190–3231, 2006.

[40] H. Bustince, "Indicator of inclusion grade for interval-valued fuzzy sets. application to approximate reasoning based on interval-valued fuzzy sets," *Int. J. of Approximate Reasoning*, vol. 23, no. 3, pp. 137–209, 2000.

[41] W. Zeng and P. Guo, "Normalized distance, similarity measure, inclusion measure and entropy of interval-valued fuzzy sets and their relationship," *Information Sciences*, vol. 178, no. 5, pp. 1334–1342, 2008.

[42] I. K. Vlachos and G. D. Sergiadis, "Subsethood, entropy, and cardinality for interval-valued fuzzy setsan algebraic derivation," *Fuzzy Sets and Systems*, vol. 158, no. 12, pp. 1384–1396, 2007.

[43] Z. Takáč, M. Minárová, J. Montero, E. Barrenechea, J. Fernandez, and H. Bustince, "Interval-valued fuzzy strong s-subsethood measures, interval-entropy and p-interval-entropy," *Information Sciences*, vol. 432, pp. 97–115, 2018.

[44] P. Grzegorzewski, "On possible and necessary inclusion of intuitionistic fuzzy sets," *Information Sciences*, vol. 181, no. 2, pp. 342–350, 2011.

[45] Z. Takáč, "Inclusion and subsethood measure for interval-valued fuzzy sets and for continuous type-2 fuzzy sets," *Fuzzy Sets and Systems*, vol. 244, no. 1, pp. 106–120, 2013.

[46] W. Zeng and H. Li, "Inclusion measures, similarity measures, and the fuzziness of fuzzy sets and their relations," *Int. J. Intelligent Systems*, vol. 21, no. 6, pp. 639–653, 2006.

[47] Y. Li, K. Qin, and X. He, "Relations among similarity measure, subsethood measure and fuzzy entropy," *Int. J. Computational Intelligence Systems*, vol. 6, no. 3, pp. 411–422, 2013.

[48] L. Xuecheng, "Entropy, distance measure and similarity measure of fuzzy sets and their relations," *Fuzzy Sets and Systems*, vol. 52, no. 3, pp. 305–318, 1992.

[49] C. P. Pappis and N. I. Karacapilidis, "A comparative assessment of measures of similarity of fuzzy values," *Fuzzy Sets and Systems*, vol. 56, no. 2, pp. 171–174, 1993.

[50] H. Bustince, E. Barrenechea, and M. Pagola, "Image thresholding using restricted equivalence functions and maximizing the measures of similarity," *Fuzzy Sets and Systems*, vol. 158, no. 5, pp. 496–516, 2007.

[51] ——, "Relationship between restricted dissimilarity functions, restricted equivalence functions and normal E_N-functions: Image thresholding invariant," *Pattern Recognition Letters*, vol. 29, no. 4, pp. 525–536, 2008.

[52] ——, "Restricted equivalence functions," *Fuzzy Sets and Systems*, vol. 157, no. 17, pp. 2333–2346, 2006.

[53] D. M. Gabbay and J. Woods, *Handbook of the History of Logic: The Many Valued and Nonmonotonic Turn in Logic*. Elsevier, 2007.

[54] R. E. Tulloss, "Assessment of similarity indices for undesirable properties and a new tripartite similarity index based on cost functions," *Mycology in Sustainable Development: Expanding Concepts, Vanishing Borders*, pp. 122–143, 1997.

[55] U. K. Met Office, "2016 Weather Summaries," accessed 2017-12-16. [Online]. Available: https://www.metoffice.gov.uk/climate/uk/summaries/2016

Shaily Kabir (S'17) received her B.Sc. and M.S. degrees in computer science and engineering (CSE) from the University of Dhaka, Bangladesh. Later, she received another Master degree in computer science (research) from Concordia University, Canada. She is a Commonwealth Scholar (2015), and currently working toward the PhD degree in the School of Computer Science, University of Nottingham, UK. She won the IEEE computational intelligence society (CIS) graduate research grant in 2018. Her main research interests include robust comparison of uncertain data received from multiple data sources and their fusion. She is a student member of the IEEE.

Christian Wagner (SM'13) received his Ph.D. in Computer Science in 2009 from the University of Essex, UK. He is a Professor of Computer Science at the University of Nottingham, and founding director of the Lab for Uncertainty in Data and Decision Making (LUCID). He has published over 100 peer-reviewed articles, focusing on modelling and handling of uncertain data arising from heterogeneous data sources (e.g., stakeholders), with a particular emphasis on designing interpretable AI based decision support systems. In 2017, he was recognised as a RISE (Recognising Inspirational Scientists and Engineers) Connector by EPSRC, UK. His work ranges from decision support in cybersecurity and environmental management to personalisation and control in manufacturing.

Dr Wagner is an Associate Editor of the IEEE Transactions on Fuzzy Systems journal, Chair of the IEEE CIS Technical Committee on Fuzzy Systems; elected member-at-large of the IEEE Computational Intelligence Society (CIS) Administrative Committee for 2018-2020. He was Special Sessions Chair at Fuzz-IEEE 2019 in New Orleans and is General Co-Chair of Fuzz-IEEE 2021 in Luxembourg.

Timothy C. Havens (SM'10) received the Ph.D. degree in electrical and computer engineering from the University of Missouri, Columbia, MO, USA, in 2010. He is currently the William and Gloria Jackson Associate Professor of Computer Systems and Associate Dean for Research in the College of Computing, Michigan Technological University (MTU), Houghton, MI, USA. He is the Director of the Institute of Computing and Cybersystems and ICC Center for Data Sciences, and co-founded the East Hancock Institute for Feline Companions.

Dr. Havens was a recipient of the Best Paper Award at the 2012 IEEE International Conference on Fuzzy Systems, the IEEE Franklin V. Taylor Award for Best Paper at the 2011 IEEE Conference on Systems, Man, and Cybernetics, and the Best Student Paper Award from the Western Journal of Nursing Research in 2009. He was the General Co-Chair of 2019 IEEE International Conference on Fuzzy Systems and is an Associate Editor of the IEEE Transactions on Fuzzy Systems.

Derek T. Anderson (SM'13) received the Ph.D. degree in electrical and computer engineering (ECE) from the University of Missouri, Columbia, MO, USA, in 2010. He is an Associate Professor in electrical engineering and computer science (EECS) at the University of Missouri and director of the Mizzou Information and Data Fusion Laboratory (MINDFUL). His research is information fusion in computational intelligence for signal/image processing, computer vision, and geospatial applications.

Dr. Anderson coauthored the 2013 best student paper in Automatic Target Recognition at SPIE, he received the best overall paper award at the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), in 2012, and the 2008 FUZZ-IEEE best student paper award. He was also the Program Co-Chair of FUZZ-IEEE 2019, Associate Editor for the IEEE Transactions on Fuzzy Systems, and an Area Editor for the International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.